DUKE MATHEMATICAL JOURNAL

Volume 9, Number 2

JUNE, 1942

DUKE UNIVERSITY PRESS

DURHAM, N. C.

DUKE MATHEMATICAL JOURNAL

This periodical is published quarterly under the auspices of Duke University by Duke University Press at Durham, North Carolina. It is printed at Mt. Royal and Guilford Avenues, Baltimore, Maryland, by the Waverly Press.

Entered as second class matter at the Post Office, Durham, North Carolina. Additional entry at the Post Office, Baltimore, Maryland.

The subscription price for the current year is four dollars, postpaid; back volumes, five dollars each, carriage extra. Subscriptions, orders for back numbers, and notice of change of address should be sent to Duke University Press, Durham, North Carolina.

Individual and institutional members of the Mathematical Association of America may subscribe to the current volume at half price. To get the reduced price, orders for subscriptions must bear the mention "Member MAA." If an order at the reduced price is placed through an agent, the purchaser must pay any commission charge incurred.

Since 1935 the Mathematical Association of America has given the Duke Mathematical Journal an annual subsidy, in return for which the half-rate has been allowed. Having served its purpose of aiding the establishment of the Journal, the subsidy is to be discontinued at the end of 1942. In view of the help already received from the Association, however, Duke University Press will for at least five years continue to allow the half-rate to any one who in 1942 is a subscriber at the reduced rate provided his subscription to the Journal and membership in the Association remain unbroken. This arrangement is expected to be permanent, but the Press reserves the right to modify or withdraw it after the five years and to change the basic rate at the beginning of any calendar year.

Manuscripts and editorial correspondence should be addressed to Duke Mathematical Journal, 4785 Duke Station, Durham, North Carolina. An Author's Manual containing detailed information about the preparation of papers for publication will be sent on request.

Authors are entitled to one hundred free reprints. Additional copies will be supplied at cost. All reprints will be furnished with covers unless the contrary is specifically requested.

The American Mathematical Society is officially represented on the Editorial Board by Professors Murray and Ward.

Made in United States of America

WAVERLY PRESS, INC.
BALTIMORE U. S. A.

THE ASYMPTOTIC FORMS OF THE SOLUTIONS OF AN ORDINARY LINEAR MATRIC DIFFERENTIAL EQUATION IN THE COMPLEX DOMAIN

By Homer E. Newell, Jr.

1. **Introduction.** The matric differential equation

$$\frac{d}{dx} Y(x, \lambda) = \{\lambda(\delta_{ij} r_j(x)) + (q_{ij}(x, \lambda))\} Y(x, \lambda),$$

under conditions to be given below, has solutions of the form $P(x, \lambda)E(x, \lambda)$, where $E(x, \lambda) = (\delta_{ij} \exp \{\lambda \int^x r_j(x)\, dx\})$ and $P(x, \lambda)$, analytic in x, reduces uniformly in x to the identity matrix when λ becomes infinite.

The present discussion rests directly upon, and extends, theory recently published by R. E. Langer[1] who showed that if the coefficient functions $r_j(x)$ are all analytic and bounded in a region of the complex plane and their differences $r_i(x) - r_j(x)$ $(i \neq j)$ are all bounded from zero and if, moreover, the functions $q_{ij}(x, \lambda)$ are analytic in x, bounded in x and λ and, for $|\lambda|$ large, admit either actual or asymptotic expansions in λ^{-1} with coefficients analytic and bounded in x, then a solution of the form $P(x, \lambda)E(x, \lambda)$ exists for the above differential equation in the neighborhood of any specified point of the given x region. This paper, for the most part, deals with equations in which the coefficient functions $r_j(x)$ may have poles and the differences $r_i(x) - r_j(x)$ $(i \neq j)$ may have zeros on the boundary of the x region in question. Regions of existence which abut such a pole or zero are established for a solution of the stated form $P(x, \lambda)E(x, \lambda)$. In addition, it is shown that the restriction to finite x regions, assumed in Langer's discussion, may be removed.

2. **The matric equation.** Throughout the considerations to follow, the differential equation[2]

$$\frac{d}{dx} Y(x, \lambda) = \{\lambda R(x) + Q(x, \lambda)\} Y(x, \lambda), \tag{2.1}$$

Received January 3, 1941.

[1] R. E. Langer, *The boundary problem of an ordinary linear differential system in the complex domain*, Trans. of the Am. Math. Soc., vol. 46(1939), pp. 151–162.

The author wishes to express his thanks to Professor Langer for most valuable suggestions received during the preparation of the present paper.

[2] In the notation adopted here, italic capitals without subscripts denote square matrices of order n. The operations of differentiation and integration are applied in accordance with the relations

$$\frac{d}{dx}(y_{ij}(x)) = \left(\frac{d}{dx} y_{ij}(x)\right), \qquad \int^x (y_{ij}(x))\, dx = \left(\int^x y_{ij}(x)\, dx\right)$$

in which the right members serve to define the left. Also, a matrix is said to be analytic if each element is analytic.

where[3] $R(x) = (\delta_{ij}r_j(x))$ and $Q(x, \lambda) = (q_{ij}(x, \lambda))$, will be assumed to satisfy the general conditions given herewith. In the complex plane the variables x and λ are to be permitted to range over suitable regions, the existence of which is presupposed as a basic hypothesis. The precise meaning of the term *suitable*, as used here, is contained in the following definition.

A closed region in the x plane consisting of a sector of a circle together with its boundary and a region in the λ plane are said to be suitable to the matric differential equation (2.1) *if for x and λ in them*:

(a) *$|\lambda|$ is unbounded;*

(b) *x_0 denoting the vertex of the circular sector in the x plane, the coefficient functions $r_j(x)$ $(j = 1, \cdots, n)$ are analytic except possibly for poles at x_0;*

(c) *the differences $r_i(x) - r_j(x)$ $(i, j = 1, \cdots, n;\ i \neq j)$ have zeros, if any, only at x_0; and*

(d) *the functions $q_{ij}(x, \lambda)$ $(i, j = 1, \cdots, n)$ are analytic in x except possibly for poles at x_0, are bounded in λ, and, for $|\lambda|$ large, admit either actual or asymptotic representations such that*

$$q_{ij}(x, \lambda) \sim \sum_{\nu=0}^{\infty} \lambda^{-\nu} q_{ij}^{(\nu)}(x) \qquad (i, j = 1, \cdots, n), \tag{2.2}$$

where the $q_{ij}^{(\nu)}(x)$ depend only on x, and where

$$q_{jj}^{(0)}(x) \equiv 0 \qquad (j = 1, \cdots, n). \tag{2.3}$$

Also, it will be assumed hereinafter that if in the suitable region in question the $r_j(x)$ are bounded, this will imply the existence of at least one index pair (i, j) $(i \neq j)$ for which $r_i(x) - r_j(x)$ has a zero at the point x_0 on the boundary. This will avoid a repetition here of the earlier results cited above in the introduction.

3. "Associated" and "fundamental" regions.

The concepts of "associated" and "fundamental" regions are due to Langer, who introduced them in the paper referred to. The details of the discussion involved in their definition are repeated here.

In a region of the x plane suitable to (2.1), the relations

$$\frac{d}{dx} R_j(x) \equiv r_j(x) \qquad (j = 1, \cdots, n) \tag{3.1}$$

can be satisfied by a set of functions $R_j(x)$ which are analytic except possibly for poles and logarithmic infinities on the boundary. Suppose such a set of functions has been chosen. The transformations

$$\xi^{ij} = \lambda\{R_i(x) - R_j(x)\} \qquad (i, j = 1, \cdots, n;\ i \neq j) \tag{3.2}$$

may then be defined, λ varying within a suitable λ region; and for each λ the

[3] The symbol δ_{ij} is the Kronecker delta.

relations (3.2) corresponding to a given index pair (i, j) will map any closed subregion[4] X of the suitable x region onto a closed region Ξ^{ij} of the complex ξ^{ij} plane.

Let λ be fixed for the moment, and consider the possibility that such a subregion X contains a set of points x_*^{ij} $(i, j = 1, \cdots, n; i \neq j)$, not necessarily distinct, such that each of the points ξ_*^{ij} obtained by the relations

$$\xi_*^{ij} = \lambda\{R_i(x_*^{ij}) - R_j(x_*^{ij})\} \qquad (i, j = 1, \cdots, n; i \neq j) \tag{3.3}$$

can be connected with every point of the respective region Ξ^{ij} by a path which lies wholly in Ξ^{ij} and along which, in passing from ξ_*^{ij} to the point ξ_{ij} in question, the abscissa is non-increasing. The existence of such a set of points plainly depends upon the shape of X.

If λ now varies, the shapes of the Ξ^{ij} do not change, as is clear from (3.2). Variations in $|\lambda|$ alone merely alter the size of any Ξ^{ij} and cannot, therefore, affect either the existence or the position of a point x_*^{ij}. But a change in arg λ causes the regions Ξ^{ij} to undergo rotations in their respective planes and an existing set of points x_*^{ij} may lose its characteristic properties thereby. This does not necessarily happen, however, and it will be shown later that such a set of points x_*^{ij} may exist independently of λ, retaining its properties for all variations in arg λ over some definite range. In reference to this possibility, the following terminology will be used.

A closed subregion X *of a suitable x region and a subregion* Λ *of a suitable* λ *region will be termed "associated" regions if there exists in* X *a set of points* x_*^{ij} $(i, j = 1, \cdots, n; i \neq j)$ *not necessarily distinct, but fixed as to* λ, *having the properties described above, and retaining them for all* λ *in the region* Λ.[5]

It may happen that for a given suitable λ region no associated region exists in the x plane. However, the possibility that there are x regions which may be associated with regions completely covering[6] the suitable λ region is not excluded thereby, and such a set of x regions may have a part in common. A definition follows.

A closed region of the x plane will be designated as a fundamental region relative to a given suitable λ *region if it is included in each of a finite number of regions* X *which are associated with regions* Λ *completely covering the suitable* λ *region in question.*[7]

4. **The existence of associated and fundamental regions.** As in §2, denote by x_0 the vertex of a circular sector suitable to (2.1). Then the following theorem can be proved.

[4] "Closed subregion", i.e., a subregion together with its boundary.

[5] Op. cit., p. 156.

[6] It is admissible that the points λ for which arg λ possesses one of finitely many specific values be regarded as "covered" by virtue of being a boundary point of a closed associated region Λ.

[7] Op. cit., p. 156.

THEOREM 1.[8] *If every $r_i(x) - r_j(x)$ $(i \neq j)$ having at x_0 a pole of order 1 is of the form $a^{ij}(x - x_0)^{-1}$, where $a^{ij}(\neq 0)$ is a constant, and if λ_0 is an arbitrarily chosen point within a suitable λ region, there exist associated regions containing x_0 and λ_0, and there exist regions containing x_0 which are fundamental relative to the suitable λ region.*

Let α be the smallest positive integer for which the functions

$$(4.1)\qquad (x - x_0)^\alpha\{r_i(x) - r_j(x)\}, \qquad (x - x_0)^\alpha\{r_i(x) - r_j(x)\}^{-1} \qquad (i, j = 1, \cdots, n; i \neq j)$$

are all bounded as $x \to x_0$. Denote by Σ any sector of which that part in which $|x - x_0|$ is sufficiently small lies in the given suitable x region, and which is of the form

$$(4.2)\qquad \beta_1 \leqq \arg(x - x_0) \leqq \beta_2,$$

where

$$(4.3)\qquad 0 < \beta_2 - \beta_1 < \pi/(\alpha + 1)(\nu + 1) \qquad (\nu = n(n - 1)/2).$$

Finally, let β_μ^{ij} $(\mu = 1, 2; i, j = 1, \cdots, n; i \neq j)$ be the limiting value of

$$\arg\{r_i(x) - r_j(x)\} + \arg \lambda_0$$

as $x \to x_0$ along $\arg(x - x_0) = \beta_\mu$.

As may be seen from (4.3) and the definition of α, the following relations hold:

$$(4.4)\qquad 0 \leqq |\beta_2^{ij} - \beta_1^{ij}| < \alpha\pi/(\alpha + 1)(\nu + 1) \qquad (i, j = 1, \cdots, n; i \neq j).$$

If γ_μ^{ij} $(\mu = 1, 2; i, j = 1, \cdots, n; i \neq j)$ are any real numbers which satisfy the conditions

$$(4.5a)\qquad \gamma_1^{ij} < \beta_\mu^{ij} < \gamma_2^{ij} \qquad (\mu = 1, 2),$$

$$(4.5b)\qquad \gamma_2^{ij} - \gamma_1^{ij} < \alpha\pi/(\alpha + 1)(\nu + 1),$$

$$(4.5c)\qquad \gamma_\mu^{ij} \equiv \gamma_\mu^{ji} \qquad (\mathrm{mod}\ \pi), (\mu = 1, 2),^{9}$$

then there exists a positive number δ such that in that part of Σ in which $0 < |x - x_0| \leqq \delta$, the following inequalities hold:

$$(4.6)\qquad \gamma_1^{ij} < \arg\{r_i(x) - r_j(x)\} + \arg\lambda_0 < \gamma_2^{ij} \qquad (i, j = 1, \cdots, n; i \neq j).$$

Let θ be a number[10] satisfying all of the inequalities

$$(4.7)\qquad (\beta_2 + \gamma_2^{ij}) - (\beta_1 + \gamma_1^{ij}) < \theta < \pi/(\nu + 1) \qquad (i, j = 1, \cdots, n; i \neq j),$$

and let τ be a number satisfying the inequality

[8] Op. cit., §4.

[9] This is consistent with the two preceding conditions, since from the relations $\{r_i(x) - r_j(x)\} = -\{r_j(x) - r_i(x)\}$, it follows that $\beta_\mu^{ij} \equiv \beta_\mu^{ji}$ (mod π), $(\mu = 1, 2)$.

[10] That such a number exists is clear from the relations (4.3) and (4.5b).

(4.8) $$\beta_1 + k\theta \leqq \tau \leqq \beta_2 + k\theta,$$

where k is some integer from the set $(0, 1, \cdots, \nu)$. Then, setting

(4.9) $$\tau^{ij} = \tau + \arg\{r_i(x) - r_j(x)\} + \arg \lambda_0,$$

the inequalities

(4.10) $$\beta_1 + \gamma_1^{ij} + k\theta < \tau^{ij} < \beta_2 + \gamma_2^{ij} + k\theta \qquad (i, j = 1, \cdots, n;\ i \neq j)$$

are satisfied.

For at least one of the admitted values of k, for all index pairs (i, j) $(i \neq j)$, τ^{ij}, satisfying (4.10), is bounded (mod π) from $\frac{1}{2}\pi$. For, if this were not so, then, since there are $\nu + 1$ values for k and, for each k, only ν essentially distinct inequalities (4.10) to consider,[11] for at least one index pair (i, j), there would be two distinct values k_1 and k_2 for which τ^{ij} satisfying the corresponding inequality (4.10) would not be bounded from $\frac{1}{2}\pi$ (mod π). But, from the left of (4.7) it would follow that

$$(k_2 + 1)\theta - k_1\theta \geqq \pi,$$

where k_1 is taken to be the smaller of k_1 and k_2. Since $k_2 + 1 - k_1 \leqq \nu + 1$, this yields

$$(\nu + 1)\theta \geqq \pi,$$

which contradicts the right hand inequality of (4.7).

Now in (4.8) take k as that value (or one of the values) for which τ^{ij}, for all pairs (i, j) $(i \neq j)$, is bounded from $\frac{1}{2}\pi$ (mod π) and observe that if τ represents the inclination of a curve in Σ (with $0 < |x - x_0| \leqq \delta$), then, for each index pair (i, j) $(i \neq j)$, τ^{ij} is the inclination of the corresponding curve in the respective ξ^{ij} plane.

If $k = 0$, consider a parallelogram X in Σ, with x_0 as one vertex, with two sides along the inclinations β_1 and β_2, and such that $|x - x_0| \leqq \delta$ in X. The boundary of such a region has an inclination τ satisfying the inequality

$$\beta_1 \leqq \tau \leqq \beta_2;$$

hence the inclinations of the boundaries of the regions Ξ^{ij} onto which X is mapped by the relations (3.2) (with $\lambda = \lambda_0$) are bounded (mod π) from $\frac{1}{2}\pi$. In addition, the boundary of every such Ξ^{ij} is cut in at most two points by each vertical line $\Re\{\xi^{ij}\}$ = constant in the respective ξ^{ij} plane, as is evident from the shape of X and the character of the map which is conformal except at the point x_0, the possibility of a reentrant angle at $\xi^{ij}(x_0)$ being excluded by (4.3). Hence, there exists in X a set of points x_*^{ij} which are independent of λ in some region Λ containing λ_0 and included in the given suitable λ region, i.e., X and Λ form a pair of associated regions.

[11] Cf. (4.5c). The inequality (4.10) corresponding to the index pair (i, j) is the same (mod π) as that corresponding to the pair (j, i).

If $k \neq 0$, there are two mutually exclusive cases. Either (1) all $\beta_\mu + \beta_\mu^{ij}$ ($\mu = 1, 2; i, j = 1, \cdots, n; i \neq j$) are not $\frac{1}{2}\pi$ (mod π) or (2) some $\beta_\mu + \beta_\mu^{ij}$ are $\frac{1}{2}\pi$ (mod π).

In the first case, there is a positive number $\delta_1 \leqq \delta$ such that the curves in the various ξ^{ij} planes which correspond under (3.2) (with $\lambda = \lambda_0$) to the parts of the rays bounding Σ on which $| x - x_0 | \leqq \delta_1$ have inclinations bounded (mod π) from $\frac{1}{2}\pi$. If the rays bounding Σ be cut at positive distances less than δ_1 from x_0 by a straight line of inclination τ satisfying (4.8), the triangle X, with vertex at x_0, so formed in Σ, can be associated (as in the case of the parallelogram above) with a region Λ containing λ_0.

In the second case, let it be recalled that each $r_i(x) - r_j(x)$ ($i \neq j$) is of the form

$$(x - x_0)^{\alpha^{ij}} \{a_0^{ij} + a_1^{ij}(x - x_0) + \cdots \} \qquad (a_0^{ij} \neq 0),$$

where the α^{ij} are integers; whence

$$\beta_\mu + \beta_\mu^{ij} = (\alpha^{ij} + 1)\beta_\mu + \arg \lambda_0 + \lim_{\substack{x \to x_0 \\ \arg (x - x_0) = \beta_\mu}} \arg \{a_0^{ij} + a_1^{ij}(x - x_0) + \cdots \} \qquad (\mu = 1, 2).$$

First consider the case in which, for all (i, j) $(i \neq j)$, $\alpha^{ij} \neq -1$. Altering Σ to Σ' by replacing β_μ by $\beta_\mu + (-1)^{\mu - 1}\Delta$ ($\mu = 1, 2$) changes $\beta_\mu + \beta_\mu^{ij}$ by the amount $(-1)^{\mu-1}(\alpha^{ij} + 1)\Delta$. Thus, by choosing Δ positive and sufficiently small, the quantities $\beta_\mu + \beta_\mu^{ij}$ corresponding to Σ' will differ from $\frac{1}{2}\pi$ (mod π) for all (i, j) $(i \neq j)$. Hence, a portion of Σ' in which $| x - x_0 |$ is sufficiently small can be associated with a λ region containing λ_0.

Next, suppose that all $\alpha^{ij} = -1$, $a_\nu^{ij} = 0$ ($\nu \geqq 1$).[12] In this case, the rays bounding Σ map into straight lines in the various ξ^{ij} planes, from which it is readily deduced that that part of Σ cut off by a straight line passing sufficiently close to x_0 and of inclination τ satisfying (4.8) can be associated with a closed region Λ containing λ_0 as a boundary point.

Finally, if the restrictions of the preceding two paragraphs do not hold so that some α^{ij} are and some are not -1, it is now clear that there is a sector Σ' differing by an arbitrarily small amount from Σ such that a portion of Σ' in which $| x - x_0 |$ is sufficiently small can be associated with a region Λ containing λ_0.

It remains to establish the existence of a fundamental region containing x_0. To this end, observe that if Λ and X are associated regions of the λ and x planes respectively, then any subregion of the suitable λ region having the same range of variation of arg λ as does Λ can also be associated with X. Also, the immediately preceding discussion has shown that every point of the suitable λ region is contained in a region associated with a region of the x plane which may be assumed to contain that part of the sector

[12] Recall the assumption made in the statement of Theorem 1.

$$\beta_1 + \frac{\beta_2 - \beta_1}{4} \leqq \arg(x - x_0) \leqq \beta_2 - \frac{\beta_2 - \beta_1}{4}$$

in which $| x - x_0 |$ is sufficiently small. Hereby the total closed range of variation of arg λ in the given suitable λ region is completely covered by subintervals corresponding to the associated subregions. But then a finite number of the λ subregions possess intervals of variation of arg λ which together completely cover the total range in the given suitable region. The corresponding finitely many associated regions of the x plane can be associated with regions entirely covering the suitable λ region, and these regions have in common a region containing x_0.

The usefulness of associated regions, as will be seen, depends upon the definitive properties of the points x_*^{ij}, together with the possibility of choosing the regions as described in the corollary immediately below. In a region X associated with a region of the λ plane, each point of an existing set x_*^{ij} can be joined to any point x of X by a path on the image of which in the respective ξ^{ij} plane in passing from ξ_*^{ij} to $\xi^{ij}(x)$ the abscissa is monotone decreasing.

Corollary to Theorem 1. *Under the conditions of Theorem* 1, *there exist associated regions with the property that the paths in the x plane referred to in the preceding paragraph may be chosen to have the further property of being bounded in length collectively, for all index pairs (i, j), by some positive constant L_0 which is independent of* λ.

In fact the associated regions exhibited in the proof of Theorem 1 possess such a property. Observe that the regions X, constructed in the proof referred to, map, for $\lambda = \lambda_0$, into regions Ξ^{ij} in each of which every point $\xi^{ij}(x)$ can be connected with ξ_*^{ij} by a path consisting entirely of rectilinear segments parallel to the axis of reals together with appropriately chosen portions of the boundary of the region Ξ^{ij} in question. Let the images in X of these paths be taken not only for $\lambda = \lambda_0$ but also for all values of λ in the associated region Λ. It is quickly verified that such a choice is appropriate. First, the path so chosen to connect a specific x_*^{ij} with a specific point x maps into a curve in the region Ξ^{ij} connecting ξ_*^{ij} with $\xi^{ij}(x)$ and along which in passing from ξ_*^{ij} to $\xi^{ij}(x)$ the abscissa is monotone decreasing. Secondly, the paths so chosen in X are independent of λ. Thirdly, these paths are bounded in length. The last statement follows from the fact that the boundary of X is of finite length, from the fact that the mapping of each Ξ^{ij} $(\lambda = \lambda_0)$ onto X is analytic in the interior and on the boundary, it being a simple matter to show that such a map carries all rectilinear paths into curves of bounded length, and from the fact that there are finitely many regions Ξ^{ij}.

5. **The solutions of the differential equation.** Let η be the smallest integer for which the functions

$$(x - x_0)^{\eta}\{r_i(x) - r_j(x)\}^{-1} \qquad (i, j = 1, \cdots, n; i \neq j) \tag{5.1}$$

are all bounded.

THEOREM 2. *Let* X *and* Λ *be a pair of associated regions having the properties described in the corollary to Theorem* 1. *Then a set of conditions sufficient that for* x *in* X, *and for all* λ *of sufficiently large absolute value*[13] *in* Λ, *there exists a solution for* (2.1) *of the form*

$$Y(x, \lambda) = \left\{I + \sum_{\nu=1}^{k-1} \lambda^{-\nu} P^{(\nu)}(x) + \lambda^{-k} B_k(x, \lambda)\right\} E(x, \lambda), \tag{5.2}$$

where k *is a positive integer, the matrices* $P^{(\nu)}(x)$ *and* $B_k(x, \lambda)$ *are analytic and bounded in* x, $B_k(x, \lambda)$ *is bounded in* λ, *and* $E(x, \lambda) = (\delta_{ij} \exp \{\lambda R_j(x)\})$, *is the following*:

(a) *the functions* $r_j(x)$ $(j = 1, \cdots, n)$ *have poles, if any, only at* x_0 ; *and the differences* $r_i(x) - r_j(x)$ $(i \neq j)$ *have zeros, if any, only at* x_0 ;

(b) *the matrix* $Q(x, \lambda)$ *is analytic in* x;

(c) *the matrices* $Q^{(\nu)}(x)$ $(\nu = 0, 1, \cdots)$ *are all analytic; and if* $\eta > 0$, *the matrices*

$$(x - x_0)^{-(k-\nu)\eta-(k-1-\nu)} Q^{(\nu)}(x) \qquad (\nu = 0, 1, \cdots, k-1)$$

are all bounded.

The formulas[14]

$$\begin{aligned} p_{ij}^{(0)} &\equiv \delta_{ij} && (i, j = 1, \cdots, n), \\ p_{ij}^{(\nu)}(x) &\equiv \{r_i(x) - r_j(x)\}^{-1}\left\{p_{ij}^{\prime(\nu-1)}(x) - \sum_{h=0}^{\nu-1}\sum_{l=1}^{n} q_{il}^{(\nu-h-1)}(x) p_{lj}^{(h)}(x)\right\} && (i \neq j), \\ p_{ii}^{\prime(\nu)}(x) &\equiv \sum_{h=0}^{\nu}\sum_{l=1}^{n} q_{il}^{(\nu-h)}(x) p_{li}^{(h)}(x) && (\nu > 0), \end{aligned} \tag{5.3}$$

with any choice of constants of integration, define in succession, for $\nu = 0, 1, \cdots, k$, a set of matrices $P^{(\nu)}(x)$ which satisfy the equations

$$P^{(\nu+1)} R - R P^{(\nu+1)} + P^{\prime(\nu)} - \sum_{h=0}^{\nu} Q^{(\nu-h)} P^{(h)} = 0 \qquad (\nu = 0, \cdots, k-1) \tag{5.4}$$

and the matrices of at least one such set will be analytic in X. This is apparent when $\eta \leqq 0$. If $\eta > 0$, let the $P^{(\nu)}(x)$ be constructed so that the elements of the matrices have zeros of the highest possible order at x_0. From hypothesis (c) of the theorem, it is readily verified that the elements of a $P^{(\nu)}(x)$ so constructed, where ν is one of the integers $1, 2, \cdots, k$, are analytic, having zeros at x_0 of order at least $(k - \nu)(\eta + 1)$.

With a set of analytic matrices $P^{(\nu)}(x)$ $(\nu = 0, \cdots, k)$, form the matrix

$$P_k(x, \lambda) \equiv \sum_{\nu=0}^{k} \lambda^{-\nu} P^{(\nu)}(x). \tag{5.5}$$

[13] Henceforth the notation "$|\lambda| > N$" will be used to replace the phrase "λ of sufficiently large absolute value".

[14] Op. cit., §5. These formulas and their use in the construction of equation (5.7) are adapted from Langer's treatment.

Then define $S(x, \lambda, k)$ by the relation

$$S(x, \lambda, k) \equiv P_k(x, \lambda)E(x, \lambda). \tag{5.6}$$

It will be found that S satisfies the equation

$$\frac{d}{dx} S = \{\lambda R + Q + \lambda^{-k} A\} S \tag{5.7}$$

in which, P_k being non-singular for $| \lambda |$ large,

$$A \sim \left\{ P'^{(k)} - \sum_{h=k}^{\infty} \lambda^{-h+k} \sum_{\nu=0}^{k} Q^{(h-\nu)} P^{(\nu)} \right\} P_k^{-1}. \tag{5.8}$$

It is evident from (5.7) and (5.8) that the elements $a_{ij}(x, \lambda)$ of the matrix $A(x, \lambda)$ are analytic in x and are bounded in x and λ. Thus, in an obvious sense, (5.7) approximates (2.1) for $| \lambda |$ large.

Now it is readily shown that if $Y(x, \lambda)$ is a solution of

$$\frac{d}{dx} \{S^{-1} Y\} = -\lambda^{-k} S^{-1} A Y, \tag{5.9}$$

then it is also a solution of (2.1). Upon integration, this becomes

$$Y(x, \lambda) = S(x) - \lambda^{-k} \int^{x} S(x) S^{-1}(\zeta) A(\zeta) Y(\zeta)\, d\zeta. \tag{5.10}$$

Setting

$$S(x) S^{-1}(\zeta) A(\zeta) = K(x, \zeta) \tag{5.11}$$

in (5.10) and then iterating, one is led to the series

$$S(x) - \lambda^{-k} S^{(1)}(x) + \lambda^{-2k} S^{(2)}(x) - \cdots + (-1)^{\nu} \lambda^{-\nu k} S^{(\nu)}(x) + \cdots, \tag{5.12}$$

where

$$S^{(\nu)}(x) = \int^{x} K(x, \zeta) S^{(\nu-1)}(\zeta)\, d\zeta \qquad (\nu = 1, 2, \cdots) \tag{5.13}$$

with $S^{(0)}(x) = S(x)$.

Now put $B^{(0)}(x) = I$ and define matrices $B^{(\nu)}(x)$ for $\nu \geqq 1$ by the equations

$$B^{(\nu)}(x) = \int^{x} K(x, \zeta) B^{(\nu-1)}(\zeta) S(\zeta) S^{-1}(x)\, d\zeta. \tag{5.14}$$

Then, for all $\nu \geqq 0$, $S^{(\nu)}(x) = B^{(\nu)}(x) S(x)$, and (5.12) takes the form

$$\{I - \lambda^{-k} B^{(1)}(x) + \lambda^{-2k} B^{(2)}(x) - \cdots \} S(x). \tag{5.15}$$

Making use of the matrices $I_{hl} = (\delta_{ih} \delta_{lj})$ and recalling (5.11), (5.14) can be written in the form

$$B^{(\nu)}(x) = \sum_{h,l=1}^{n} \int_{x^{hl}}^{x} S(x) I_{hh} S^{-1}(\zeta) A(\zeta) B^{(\nu-1)}(\zeta) S(\zeta) I_{ll} S^{-1}(x)\, d\zeta,$$

where, for the moment, the points x^{hl} are formally undetermined.

If we note that

$$S(x)I_{hh}S^{-1}(\zeta) = P_k(x)I_{hh}P_k^{-1}(\zeta)e^{\lambda R_h(x)-\lambda R_h(\zeta)},$$

this becomes[15]

$$(5.16)\quad B^{(\nu)}(x) = \sum_{h,l=1}^{n}\int_{x^{hl}}^{x} P_k(x)I_{hh}P_k^{-1}(\zeta)A(\zeta)B^{(\nu-1)}(\zeta)P_k(\zeta)I_{ll}P_k^{-1}(x)e^{\xi^{hl}(x)-\xi^{hl}(\zeta)}\,d\zeta.$$

Now set[16] $x^{hl} = x_*^{hl}$ and choose the paths of integration to be bounded in length uniformly in λ by a positive constant L_0 and such that on those corresponding to the index pair (h, l), $\Re\{\xi^{hl}(x) - \xi^{hl}(\zeta)\} \leqq 0$. The possibility of satisfying the latter requirement is an immediate consequence of the definitive properties of the points x_*^{hl} $(h \neq l)$. Then along the respective paths of integration the following relations hold:

$$|e^{\xi^{hl}(x)-\xi^{hl}(\zeta)}| \leqq 1 \qquad (h, l = 1, \cdots, n).$$

Thus (the matrices A, P_k, P_k^{-1} being bounded in x, and in λ for $|\lambda|$ large), if there exists a sequence of positive numbers $b^{(\nu)}$ $(\nu = 0, 1, \cdots)$ such that for all ν:

$$|b_{ij}^{(\nu)}| \leqq b^{(\nu)} \qquad (i, j = 1, \cdots, n),$$

it follows directly that, for some positive m, and for $|\lambda|$ large,

$$|b_{ij}^{(\nu)}| \leqq mb^{(\nu-1)} \qquad (i, j = 1, \cdots, n; \nu \geqq 1).$$

But $b^{(0)}$ can be taken as 1, whence by induction:

$$(5.17)\qquad |b_{ij}^{(\nu)}| \leqq m^{\nu} = b^{(\nu)} \qquad (i, j = 1, \cdots, n; \nu \geqq 0).$$

Hence the series $\sum (-1)^{\nu}\lambda^{-\nu k}B^{(\nu)}$ is dominated by the series $\sum \lambda^{-\nu k}(m^{\nu})$ and, therefore, converges absolutely and uniformly in x and λ for x in X and $|\lambda| > N$. Thus a function $Y(x, \lambda)$ is defined in X by the relation

$$(5.18)\qquad Y(x, \lambda) = \sum_{\nu=0}^{\infty} (-1)^{\nu}\lambda^{-\nu k}B^{(\nu)}(x)S(x) = U(x, \lambda)S(x, \lambda).$$

The function $Y(x, \lambda)$ so defined is a solution of (2.1). To show this, multiply (5.18) on the left by $S^{-1}(x)$, and differentiate.

$$\{S^{-1}Y\}' = (S^{-1})'US + S^{-1}U'S + S^{-1}US'.$$

Since the series $U(x, \lambda)$ is uniformly convergent in x, it may be differentiated term by term. Simplifying, the above reduces to (5.9), whence $Y(x, \lambda)$ is a solution of (2.1).

Finally, factoring out $E(x, \lambda)$, (5.18) takes the form (5.2).

[15] The notation ξ^{hl} is undefined for $h = l$. Let it be given the interpretation: $\xi^{hh} \equiv 0$ $(h = 1, \cdots, n)$.

[16] The points x_*^{hl} have been defined only for $h \neq l$. Let them be chosen arbitrarily when $h = l$.

COROLLARY. *Under the conditions of the theorem, there exist fundamental regions containing x_0 wherein, for $|\lambda| > N$ in a suitable λ region, a solution $Y(x, \lambda)$ of* (2.1) *exists such that*

$$\lambda^{k-1}\left\{Y(x, \lambda)E^{-1}(x, \lambda) - \sum_{\nu=0}^{k-1} \lambda^{-\nu} P^{(\nu)}(x)\right\} \to 0$$

uniformly in x as $\lambda \to \infty$, where the $P^{(\nu)}(x)$ $(\nu = 0, \cdots, k-1)$ are analytic and bounded.

The corollary is an immediate consequence of the fact that from Theorem 1 and its corollary there exists a fundamental region containing x_0 which is the intersection of a finite set of x regions associated with regions covering all of the given suitable λ region and to each of which Theorem 2 applies.

6. **The case of a meromorphic** $Q(x, \lambda)$. The results of Theorem 2 and its corollary may be generalized to cases in which the elements of the matrix $Q(x, \lambda)$ have poles. Let X and Λ be a pair of associated regions possessing the properties described in the corollary to Theorem 1, and suppose that for x in X the following conditions hold:

(a) *the functions $r_i(x) - r_j(x)$ $(i, j = 1, \cdots, n; i \neq j)$ are bounded from zero, having poles at x_0 on the boundary of* X;

(b) $\eta \leqq -2$ (cf. page 251, line 4 up);

(c) *the functions $q_{ij}(x, \lambda)$ $(i, j = 1, \cdots, n)$ have poles, if any, only at x_0.*

THEOREM 3. *A solution for* (2.1) *of the form* (5.2) *with $k = 1$ exists in* X *for $|\lambda| > N$ in Λ if, in addition to conditions* (a), (b), *and* (c) *above, the following conditions also hold*:

(d) *the matrices $Q^{(\nu)}(x)$ $(\nu \geqq 1)$ are analytic;*

(e) *the matrix*

$$(x - x_0)^{-\frac{\eta-1}{3}} Q^{(0)}(x)$$

is bounded and

(f) *the functions*

$$(x - x_0)^{\frac{\eta+2}{3}} q_{jj}^{(1)}(x) \qquad (j = 1, \cdots, n)$$

are all bounded.

Given a pair of suitable regions in the x and λ planes, there exists, under the specified conditions, a fundamental region containing x_0 in which, for $|\lambda| > N$ in the suitable λ region, there is a solution $Y(x, \lambda)$ of (2.1) *such that*

$$Y(x, \lambda)E^{-1}(x, \lambda) \to I$$

uniformly in x as $\lambda \to \infty$.

The conditions of the theorem admit of the determination, from formulas (5.3), of matrices $P^{(0)}(x)$ and $P^{(1)}(x)$ which satisfy the relations (5.4) for $\nu = 0, 1$

and such that the elements of $P^{(1)}(x)$ all have zeros at x_0 of order at least $-(\eta - 1)/3$. Using matrices so determined, form the matrix

$$P_1(x, \lambda) \equiv \sum_{\nu=0}^{1} \lambda^{-\nu} P^{(\nu)}(x).$$

Then $S(x, \lambda, 1)$, defined by the relation

$$S(x, \lambda, 1) \equiv P_1(x, \lambda)E(x, \lambda),$$

satisfies the equation

$$\frac{d}{dx}S = \{\lambda R + Q + \lambda^{-1}A\}S \tag{6.1}$$

in which

$$A \sim \left\{P'^{(1)}(x) - \sum_{h=1}^{\infty} \lambda^{-h+1} \sum_{\nu=0}^{1} Q^{(h-\nu)}(x)P^{(\nu)}(x)\right\} P_1^{-1}(x, \lambda). \tag{6.2}$$

Since the elements of $P^{(1)}(x)$ have zeros at x_0 of order at least $-(\eta - 1)/3$, it follows from (6.1) and (6.2) that the elements $a_{ij}(x, \lambda)$ of $A(x, \lambda)$ are analytic in x and are bounded in x and λ.

It is now apparent that the proof of the theorem can be made precisely similar to that used to establish Theorem 2 and its corollary.

THEOREM 4. *A solution for* (2.1) *of the form* (5.2) *with* $k = 2$ *exists in* X *for* $|\lambda| > N$ *in* Λ *if, in addition to conditions* (a), (b), *and* (c), *the following conditions also hold*:

(d) *the matrices* $Q^{(\nu)}(x)$ $(\nu \geqq 2)$ *are analytic*;

(e) *the matrices*

$$(x - x_0)^{-\frac{\eta-2}{4}} Q^{(\nu)}(x) \qquad (\nu = 0, 1)$$

are bounded and

(f) *the functions*

$$(x - x_0)^{\frac{\eta+2}{2\nu}} q_{jj}^{(\nu)}(x) \qquad (\nu = 1, 2; j = 1, \cdots, n)$$

are all bounded.

Given a pair of suitable regions in the x *and* λ *planes, there is a fundamental region containing* x_0 *in which, for* $|\lambda| > N$ *in the suitable* λ *region, there exists a solution* $Y(x, \lambda)$ *of* (2.1) *such that*

$$\lambda\{Y(x, \lambda)E^{-1}(x, \lambda) - I - \lambda^{-1}P^{(1)}(x)\} \to 0$$

uniformly in x *as* $\lambda \to \infty$, *where* $P^{(1)}(x)$ *is analytic.*

It is readily verified that matrices $P^{(0)}(x)$, $P^{(1)}(x)$, and $P^{(2)}(x)$ can be constructed from (5.3) to satisfy the equations (5.4) for $\nu = 0, 1, 2$ and such that the elements of $P^{(1)}(x)$ and $P^{(2)}(x)$ have zeros at x_0 of order at least $-(\eta - 2)/4$. With matrices so constructed, form

$$P_2(x, \lambda) \equiv \sum_{\nu=0}^{2} \lambda^{-\nu} P^{(\nu)}(x).$$

Then $S(x, \lambda, 2)$, defined by the relation

$$S(x, \lambda, 2) \equiv P_2(x, \lambda)E(x, \lambda),$$

satisfies

$$\frac{d}{dx} S = \{\lambda R + Q + \lambda^{-2} A\} S$$

in which

$$A \sim \left\{P'^{(2)}(x) - \sum_{h=2}^{\infty} \lambda^{-h+2} \sum_{\nu=0}^{2} Q^{(h-\nu)}(x)P^{(\nu)}(x)\right\} P_2^{-1}(x, \lambda).$$

Since the elements of $P^{(1)}(x)$ and $P^{(2)}(x)$ have at x_0 zeros of order at least $-(\eta - 2)/4$, it can be shown that A is analytic in x and is bounded in x and λ. The proof of the theorem can now be continued as in the proofs of Theorem 2 and its corollary.

THEOREM 5. *A solution for* (2.1) *of the form* (5.2) *with k an arbitrary positive integer exists in* X *for* $|\lambda| > N$ *in* Λ *if, in addition to conditions* (a), (b), *and* (c), *the following conditions also hold*:

(d) $Q^{(\nu)}(x) \equiv 0 \qquad (\nu \geqq 1)$ *and*

(e) *the matrix*

$$(x - x_0)^{-\frac{m}{2m+1}(\eta-1)} Q^{(0)}(x)$$

is bounded, where m is some positive integer.

Given a pair of suitable regions in the x and λ planes, there exists, under the specified conditions, a fundamental region containing x_0 in which, for $|\lambda| > N$ in the suitable λ region, there is a solution $Y(x, \lambda)$ of (2.1) *such that*

$$Y(x, \lambda) \sim \left\{\sum_{\nu=0}^{\infty} \lambda^{-\nu} P^{(\nu)}(x)\right\} E(x, \lambda),$$

where the $P^{(\nu)}(x)$ are analytic.

Under the conditions given in the statement of the theorem, it is easily shown by induction that matrices $P^{(\nu)}(x)$ $(\nu = 1, 2, 3, \cdots)$ may be determined to have respectively zeros at x_0 of order at least $-\nu(\eta - 1)/(2m + 1)$. Thus the matrix A which has the expansion

$$A \sim \{P'^{(k)}(x) - Q^{(0)}(x)P^{(k)}(x)\}P_k^{-1}(x, \lambda)$$

(obtained from (5.8) by setting $Q^{(\nu)}(x) \equiv 0$ for $\nu \geqq 1$) is analytic in x and bounded in x and λ if $k \geqq m$. Hence Theorem 5 follows as in the proofs of Theorem 2 and its corollary, for $k \geqq m$, and thereby for any positive integer k.

7. **Extension to infinite regions.** The results obtained for finite x regions in

the preceding two sections under prescribed conditions on the coefficient functions of (2.1) are also valid for infinite regions under "equivalent", but not formally identical, conditions. This is seen as follows.

In (2.1), let $(x - x_0)$ be replaced by $-\zeta^{-1}$. The equation takes the form

$$\text{(2.1')} \quad \frac{d}{d\zeta} Y(x_0 - \zeta^{-1}, \lambda) = \{\lambda R(x_0 - \zeta^{-1})\zeta^{-2} + Q(x_0 - \zeta^{-1}, \lambda)\zeta^{-2}\} Y(x_0 - \zeta^{-1}, \lambda)$$

in which the new coefficient functions may be regarded as satisfying at ∞ conditions equivalent to those originally holding at the finite point x_0, in the sense that the solution $U(\zeta, \lambda) = Y(x_0 - \zeta^{-1}, \lambda)$ of (2.1') has the same properties at $\zeta = \infty$ as does $Y(x, \lambda)$ at x_0.

Since (2.1') is of the same form as (2.1), the effect of the above transformation may be reproduced qualitatively by a restatement of conditions on the coefficient functions of (2.1) so as to apply in the neighborhood of ∞. Thus, by a suitable rewording, the previous theorems may be stated so as to hold for infinite regions. For example, the theorem of Langer's paper[17] may be given the form:

THEOREM 6. *A set of conditions sufficient that there exist infinite regions* X *and* Λ *such that for* x *in* X *and for* $|\lambda| > N$ *in* Λ *there is a solution for* (2.1) *of the form*

$$Y(x, \lambda) \sim \left\{ \sum_{\nu=0}^{\infty} \lambda^{-\nu} P^{(\nu)}(x) \right\} E(x, \lambda),$$

the $P^{(\nu)}(x)$ *being analytic and bounded, is that for* x *in the neighborhood of* ∞ *and for* λ *in* Λ:

(a) *the matrix* $x^2R(x)$ *be analytic;*

(b) *the differences* $r_i(x) - r_j(x)$ $(i, j = 1, \cdots, n; i \neq j)$ *have zeros at* $x = \infty$ *of order exactly* 2;

(c) *the matrix* $x^2Q(x, \lambda)$ *be analytic in* x; *and*

(d) *for* $|\lambda| > N$:

$$Q(x, \lambda) \sim \sum_{\nu=0}^{\infty} \lambda^{-\nu} Q^{(\nu)}(x),$$

where the matrices $x^2Q^{(\nu)}(x)$ $(\nu = 0, 1, \cdots)$ *are analytic.*

The conditions given in the above theorem are also sufficient that $x = \infty$ be an interior point of the x regions in question.

UNIVERSITY OF MARYLAND.

[17] Op. cit., p. 162.

A SELF-RECIPROCAL FUNCTION

By R. S. Varma

The object of this paper is to establish the following theorem.

The function

$$f(x) = x^{\nu-p-q+\frac{1}{2}} e^{-\frac{1}{4}x^2} J_p(x) J_q(x)$$

is[1] R_ν, *provided* $\Re(\nu) > -1$.

We shall require the integral

$$I = \int_0^\infty x^{s-1} e^{-\frac{1}{4}x^2} W_{k,m}(\tfrac{1}{2}x^2) J_p(ax) J_q(ax)\, dx.$$

This can be evaluated by substituting for $J_p(ax)J_q(ax)$ the equivalent infinite series (see [4], p. 380)

$$\sum_{r=0}^\infty \frac{(-1)^r \Gamma(p+q+2r+1)(\frac{1}{2}ax)^{p+q+2r}}{r!\,\Gamma(p+r+1)\Gamma(q+r+1)\Gamma(p+q+r+1)}$$

and integrating term by term by the help of the integral (see [2])

$$\int_0^\infty x^{l-1} e^{-(\alpha^2+\frac{1}{2})x} W_{k,m}(x)\, dx$$

$$= \frac{\Gamma(l+m+\frac{1}{2})\Gamma(l-m+\frac{1}{2})}{\Gamma(l-k+1)}\, {}_2F_1(l+m+\tfrac{1}{2}, l-m+\tfrac{1}{2}; l-k+1; -\alpha^2)$$

$$(l \pm m + \tfrac{1}{2} > 0, |\Im(\alpha)| < 1).$$

We then obtain

$$I = (\tfrac{1}{2}a)^{p+q} \frac{2^{\frac{1}{2}s-1}\Gamma(\frac{1}{2}s+\frac{1}{2}p+\frac{1}{2}q+m+\frac{1}{2})\Gamma(\frac{1}{2}s+\frac{1}{2}p+\frac{1}{2}q-m+\frac{1}{2})}{\Gamma(p+1)\Gamma(q+1)\Gamma(\frac{1}{2}s+\frac{1}{2}p+\frac{1}{2}q-k+1)}$$

$$\times\, {}_4F_4\left[\begin{matrix}\frac{1}{2}p+\frac{1}{2}q+\frac{1}{2}, & \frac{1}{2}p+\frac{1}{2}q+1, & \frac{1}{2}s+\frac{1}{2}p+\frac{1}{2}q+m+\frac{1}{2}, & \frac{1}{2}s+\frac{1}{2}p+\frac{1}{2}q-m+\frac{1}{2}; \; -2a^2 \\ p+1, & q+1, & p+q+1, & \frac{1}{2}s+\frac{1}{2}p+\frac{1}{2}q-k+1\end{matrix}\right]. \tag{1}$$

Term by term integration is justified by virtue of

$$|W_{k,m}(x)| = O(e^{-\frac{1}{2}x}x^k), \qquad |J_\nu(x)| = O(x^{-\frac{1}{2}}) \tag{2}$$

and by virtue of the size of the terms in the series of $J_p(ax)J_q(ax)$. Hence the result (1) has been shown to be true when $\Re(p) > -1$, $\Re(q) > -1$, and

Received January 24, 1941.

[1] Following Hardy and Littlewood, we say that a function is R_ν when it is self-reciprocal in the Hankel-transform of order ν.

$\Re(s + p + q \pm 2m + 1) > 0$. By the theory of analytical continuation, the first two conditions $\Re(p) > -1$ and $\Re(q) > -1$ may be removed.

It may be noted that, since

$$W_{k,\pm\frac{1}{4}}(\tfrac{1}{2}x^2) = 2^{-k}x^{\frac{1}{2}}D_{2k-\frac{1}{2}}(x),$$

the above integral reduces for $m = \pm\frac{1}{4}$ to a result deduced by me (see [3]) elsewhere.

To establish the above theorem, let us assume that

$$x^{\lambda}e^{-\frac{1}{4}x^2}W_{k,m}(\tfrac{1}{2}x^2)J_p(x)J_q(x)$$

is R_ν, i.e., that

$$\begin{aligned} x^{\lambda}e^{-\frac{1}{4}x^2}W_{k,m}(\tfrac{1}{2}x^2)J_p(x)J_q(x) \\ = \int_0^\infty (xy)^{\frac{1}{2}}J_\nu(xy)y^{\lambda}e^{-\frac{1}{4}y^2}W_{k,m}(\tfrac{1}{2}y^2)J_p(y)J_q(y)\,dy. \end{aligned} \tag{3}$$

If, therefore, we set

$$f(s) = \int_0^\infty x^{s-1}e^{-\frac{1}{4}x^2}W_{k,m}(\tfrac{1}{2}x^2)J_p(x)J_q(x)\,dx,$$

then

$$\begin{aligned} f(\lambda + s) &= \int_0^\infty x^{\lambda+s-1}e^{-\frac{1}{4}x^2}W_{k,m}(\tfrac{1}{2}x^2)J_p(x)J_q(x)\,dx \\ &= \int_0^\infty x^{s-1}\,dx\int_0^\infty (xy)^{\frac{1}{2}}J_\nu(xy)y^{\lambda}e^{-\frac{1}{4}y^2}W_{k,m}(\tfrac{1}{2}y^2)J_p(y)J_q(y)\,dy \\ &= \int_0^\infty y^{\lambda+\frac{1}{2}}e^{-\frac{1}{4}y^2}W_{k,m}(\tfrac{1}{2}y^2)J_p(y)J_q(y)\,dy\int_0^\infty x^{s-\frac{1}{2}}J_\nu(xy)\,dx. \end{aligned}$$

The change in the order of integration is permissible by de la Vallé Poussin's theorem (see [1]) on account of the asymptotic estimates (2). Since (see [4], p. 383)

$$\int_0^\infty x^q J_n(ax)\,dx = \frac{2^q\Gamma(\frac{1}{2}n + \frac{1}{2}q + \frac{1}{2})}{a^{q+1}\Gamma(\frac{1}{2}n - \frac{1}{2}q + \frac{1}{2})} \qquad [\Re(q) < \tfrac{1}{2},\ \Re(n + q + 1) > 0],$$

$$\begin{aligned} f(\lambda + s) &= 2^{s-\frac{1}{2}}\frac{\Gamma(\frac{1}{2}\nu + \frac{1}{2}s + \frac{1}{4})}{\Gamma(\frac{1}{2}\nu - \frac{1}{2}s + \frac{3}{4})}\int_0^\infty y^{\lambda-s}e^{-\frac{1}{4}y^2}W_{k,m}(\tfrac{1}{2}y^2)J_p(y)J_q(y)\,dy \\ &= 2^{s-\frac{1}{2}}\frac{\Gamma(\frac{1}{2}\nu + \frac{1}{2}s + \frac{1}{4})}{\Gamma(\frac{1}{2}\nu - \frac{1}{2}s + \frac{3}{4})}f(\lambda - s + 1). \end{aligned} \tag{4}$$

Considering the equation (4) in conjunction with the result (1), we have, for the validity of (3), the equation

$$\begin{aligned} &\Gamma(\tfrac{1}{2}\lambda + \tfrac{1}{2}s + \tfrac{1}{2}p + \tfrac{1}{2}q + m + \tfrac{1}{2})\Gamma(\tfrac{1}{2}\lambda + \tfrac{1}{2}s + \tfrac{1}{2}p + \tfrac{1}{2}q - m + \tfrac{1}{2}) \\ &\qquad \times \Gamma(\tfrac{1}{2}\nu - \tfrac{1}{2}s + \tfrac{3}{4})\Gamma(\tfrac{1}{2}\lambda + \tfrac{1}{2}p + \tfrac{1}{2}q - \tfrac{1}{2}s - k + \tfrac{3}{2}) \end{aligned}$$

$$= \Gamma(\tfrac{1}{2}\lambda + \tfrac{1}{2}s + \tfrac{1}{2}p + \tfrac{1}{2}q - k + 1)\Gamma(\tfrac{1}{2}\nu + \tfrac{1}{2}s + \tfrac{1}{4})$$
$$\times \Gamma(\tfrac{1}{2}\lambda - \tfrac{1}{2}s + \tfrac{1}{2}p + \tfrac{1}{2}q + m + 1)\Gamma(\tfrac{1}{2}\lambda - \tfrac{1}{2}s + \tfrac{1}{2}p + \tfrac{1}{2}q - m + 1)$$

which is satisfied by taking

$$k = m + \tfrac{1}{2}, \qquad \lambda = \nu - 2m - p - q - \tfrac{1}{2}.$$

Hence,

$$x^{\nu-2m-p-q-\frac{1}{2}}e^{-\frac{1}{4}x^2}W_{m+\frac{1}{2},m}(\tfrac{1}{2}x^2)J_p(x)J_q(x)$$

is R_ν. Since

$$W_{m+\frac{1}{2},m}(x) = x^{m+\frac{1}{2}}e^{-\frac{1}{2}x},$$

we have, therefore, proved that

$$x^{\nu-p-q+\frac{1}{2}}e^{-\frac{1}{2}x^2}J_p(x)J_q(x)$$

is R_ν, i.e., that

$$x^{\nu-p-q+\frac{1}{2}}e^{-\frac{1}{2}x^2}J_p(x)J_q(x) = \int_0^\infty (xy)^{\frac{1}{2}}J_\nu(xy)y^{\nu-p-q+\frac{1}{2}}e^{-\frac{1}{2}y^2}J_p(y)J_q(y)\,dy,$$

provided $\Re(\nu) > -1$.

Bibliography

1. T. J. I'A. Bromwich, *Infinite Series* (Second edition, revised), London, 1926, p. 504.
2. S. Goldstein, *Operational representation of Whittaker's confluent hypergeometric function and Weber's parabolic cylinder function*, Proceedings of the London Mathematical Society (2), vol. 34(1932), pp. 103–125.
3. R. S. Varma, *Some infinite integrals involving parabolic cylinder functions*, Journal de Mathématiques, vol. 18(1939), pp. 157–166.
4. E. T. Whittaker and G. N. Watson, *Modern Analysis*, Fourth edition, Cambridge, England, 1927.

The University, Lucknow, India.

THE STRUCTURE OF THE GROUP OF $\mathfrak{P}$-ADIC 1-UNITS

By David Gilbarg

Introduction. It is well known that in local $\mathfrak{P}$-adic fields the logarithm function can be defined by the series

$$-\log(1-x) = \sum_{\nu=1}^{\infty} \frac{x^\nu}{\nu},$$

where x is a $\mathfrak{P}$-adic number, the series converging for all x with $|x|_\mathfrak{P} < 1$. Although the logarithm is needed in various connections for algebraic number theory, very little is known about it. Even its value domain is still unknown, except for this fact: $\log(1-x)$ maps the set x for which $|x|_\mathfrak{P} < |p|^{\frac{1}{p-1}}$ onto itself in a one-to-one way. However, the mapping of those x for which $|p|^{\frac{1}{p-1}} \leqq |x|_\mathfrak{P} < 1$ is still unknown, and it is this which must be investigated.

Many of the explicit formulas for the reciprocity law in algebraic number fields are best stated by means of the $\mathfrak{P}$-adic logarithm.[1] Although these explicit formulas have been proved, they are not clearly understood; it is probable that complete knowledge of the value domain of the $\mathfrak{P}$-adic logarithm would better our understanding of the formulas. This knowledge would be of use also for other applications to algebraic number theory.

Let K be a $\mathfrak{P}$-adic number field; then all units ϵ which are congruent to 1 modulo $\mathfrak{P}$—the 1-units of Hensel—constitute the set of elements in K having logarithms. If the structure of this multiplicative group of 1-units were completely known in some convenient way, then also the value domain of the $\mathfrak{P}$-adic logarithm would be known. M. Krasner has attacked this problem,[2] considering the general case where K is normal over a field k. His method was the following. Let G be the Galois group of K/k, $\sigma, \tau, \cdots$, its elements, and form the group ring Γ of G taken over the ring of p-adic integers; Γ consists of elements $\zeta = a\sigma + b\tau + \cdots$, where $a, b, \cdots$ are p-adic integers. If ϵ is a 1-unit, then the hypercomplex power ϵ^ζ can be defined in the usual way,

$$\epsilon^\zeta = (\sigma\epsilon)^a(\tau\epsilon)^b \cdots = \sigma(\epsilon^a)\tau(\epsilon^b) \cdots$$

with $(\epsilon^{\zeta_1})^{\zeta_2} = \epsilon^{\zeta_2\zeta_1}$. In this way, Γ is seen to be a ring of operators on the group of 1-units. Krasner tried to find a minimal basis for the 1-units, taking the hypercomplex exponents Γ as operator domain; that is, he tried to find the fewest number of 1-units $\epsilon_1, \epsilon_2, \cdots, \epsilon_r$, such that $\epsilon_1^\Gamma \epsilon_2^\Gamma \cdots \epsilon_r^\Gamma$ give all 1-units in K (except perhaps for roots of unity). It was hoped that an independent

Received July 28, 1941. The author wishes to express his indebtedness to Professor Artin and Dr. Whaples of the Indiana University mathematics department for their assistance in preparing this paper.

[1] See [2] (numbers in brackets refer to the bibliography); see chapter IV on explicit formulas for the reciprocity law and the $\mathfrak{P}$-adic logarithm.

[2] See M. Krasner [5]. All references to Krasner have to do with this paper.

basis could be found, namely, 1-units $\epsilon_1, \cdots, \epsilon_r$ and possibly a p^s-th root of unity η, such that $\eta^a \epsilon_1^{\zeta_1} \epsilon_2^{\zeta_2} \cdots \epsilon_r^{\zeta_r} = 1$ implies: $a \equiv 0 \bmod p^s$, $\zeta_1 = \zeta_2 = \cdots = \zeta_r = 0$. Such an independent set will be called a *normal basis*. Could a normal basis be found, then the structure of the group of 1-units in K would be completely known in terms of these basis elements and their behavior under the automorphisms of K/k.

Krasner was able to prove the following theorems:

I. If K/k is of degree prime to p, then the 1-units of K have a normal basis.

II. Let K/k be regular, that is, without primitive p-th roots of unity. Then, if K/k is without higher ramification,[3] the 1-units of K have a normal basis over k.

It was intimated that more general theorems than these would be proved at a later date. However, I shall show here that Theorems I and II cannot be included in more general simple theorems. Counterexamples will be given in very simple fields to show that the assumption on the degree of K/k in Theorem I cannot be dropped, and the converse to Theorem II will be proved.

Since Theorem I is of some interest, I offer a simplified proof of it. The method here differs entirely from that of Krasner, and is relatively easy, although at one point a theorem requiring the theory of algebras is used.

1. **Groups of 1-units which have a normal basis.** Although general normal fields K/k do not have a normal basis for 1-units, it is easy to show that every such field contains a sub-group of finite index (in the group of 1-units) which does possess a normal basis.

To see this, let H be the group of 1-units in K/k, and consider the logarithms of all elements in H. Since every 1-unit is determined, within a root of unity,[4] by its logarithm, it follows that there is an isomorphism between the additive group of logarithms and the multiplicative group of 1-units modulo the roots of unity in H:

$$\frac{\mathrm{H}}{(\zeta)} \cong \log \mathrm{H},$$

where ζ generates the roots of unity in H.

Let $G = \sigma, \tau, \cdots$ be the group of K/k. It is well known that K/k has a field basis which is normal[5] $\sigma\alpha, \tau\alpha, \cdots$, consisting of the conjugates of a single element, α. If $\beta_1, \beta_2, \cdots, \beta_n$ is a basis for k/R_p (R_p being the rational p-adic field), then clearly the elements $\sigma(\beta_i\alpha)$ are a basis for K/R_p. In particular, the integers of K are expressible by linear combinations, with coefficients in R_p, of

[3] Let $\mathfrak{p}$ and $\mathfrak{P}$ be the prime ideals in k and K respectively; then $\mathfrak{p} = \mathfrak{P}^e$, where e is the ramification index of K/k. If e is divisible by p, then K/k is said to have higher ramification.

[4] Log $x = 0$ implies x is a root of unity with order a power of p (these being the only roots of unity that are at the same time 1-units).

[5] The normal basis meant here is a field basis, and is not to be confused with a normal basis for 1-units.

the $\sigma(\beta_i\alpha)$. The denominators of these coefficients are bounded by the discriminant of the $\sigma(\beta_i\alpha)$. Consequently, on dividing each $\beta_i\alpha$ by a high enough power of p (say the discriminant), a new set of basis elements, $\sigma(\mathrm{A}_i)$, is obtained, such that every integer of K is expressible as linear combination of the $\sigma(\mathrm{A}_i)$, with coefficients among the integers $\mathfrak{o}$ of R_p. Thus the group, $\Theta = \sum_{\sigma,i} \mathfrak{o}\sigma(\mathrm{A}_i)$, contains the integers of K as subgroup, and also, therefore, the group log H, which consists only of integral elements.

Contained in log H are all those integers of K which lie in the convergence domain of the $\mathfrak{P}$-adic exponential function[6] (for, $\log e^x = x$, provided e^x converges). Now, if the basis elements $\sigma(\mathrm{A}_i)$ are multiplied by a high enough power of p, then new field basis elements, which we call $\sigma(\mathrm{B}_i)$, are obtained, and such that the entire group of elements in $\sum_{\sigma,i} \mathfrak{o}\sigma(\mathrm{B}_i)$ lies in the convergence domain of the $\mathfrak{P}$-adic exponential, and is therefore contained in log H. But, by its construction, this group, $\sum_{\sigma,i} \mathfrak{o}\sigma(\mathrm{B}_i)$, is of finite index in the group Θ. This shows two things which will be needed: (1) $\sum_{\sigma,i} \mathfrak{o}\sigma(\mathrm{B}_i)$ is of finite index in log H, (2) log H is of finite index in the group $\Theta = \sum_{\sigma,i} \mathfrak{o}\sigma(\mathrm{A}_i)$ (hence also in the integers of K).

It follows from (1) and from the isomorphism $\mathrm{H}/(\zeta) \cong \log \mathrm{H}$, that the multiplicative group, $\exp \sum_{\sigma,i} \mathfrak{o}\sigma(\mathrm{B}_i)$, obtained by taking the exponential of every element in $\sum_{\sigma,i} \mathfrak{o}\sigma(\mathrm{B}_i)$ is of finite index in H. The 1-units $\epsilon_i = e^{\mathrm{B}_i}$ form a normal basis for this subgroup. This is seen easily as follows. If Γ is the group ring generated by G over $\mathfrak{o}$, then

$$\epsilon_1^{\Gamma}\epsilon_2^{\Gamma} \cdots \epsilon_n^{\Gamma} = (e^{\mathrm{B}_1})^{\Gamma}(e^{\mathrm{B}_2})^{\Gamma} \cdots (e^{\mathrm{B}_n})^{\Gamma} = e^{\Gamma\mathrm{B}_1+\Gamma\mathrm{B}_2+\cdots+\Gamma\mathrm{B}_n}$$
$$= \exp \sum_{\sigma,i} \mathfrak{o}\sigma(\mathrm{B}_i)$$

and this representation for the elements of the subgroup is unique, since

$$1 = \epsilon_1^{\gamma_1}\epsilon_2^{\gamma_2} \cdots \epsilon_n^{\gamma_n} = e^{\gamma_1\mathrm{B}_1+\gamma_2\mathrm{B}_2+\cdots+\gamma_n\mathrm{B}_n}$$

implies that $\gamma_1 = \gamma_2 = \cdots = \gamma_n = 0$ because $e^x = 1$ only for $x = 0$ and $\sum_i \gamma_i\mathrm{B}_i = 0$ only for all $\gamma_i = 0$.

The above shows that *every* field contains a subgroup with normal basis of finite index in the entire group of 1-units. Although this is not a strong statement about the structure of the 1-units, it is sufficient for local class field theory

[6] The $\mathfrak{P}$-adic exponential is defined by the power series for e^x, the arguments being $\mathfrak{P}$-adic numbers. The convergence domain for the exponential is the set of numbers x such that $|x|_{\mathfrak{P}} < |p|^{\frac{1}{p-1}}$. For x in the convergence domain, $e^x \equiv 1 \bmod \mathfrak{P}$, so e^x is a 1-unit. A number is uniquely determined by its exponential.

where Herbrand's lemma is used for making index computation, and in the process requires knowledge only of the subgroup.[7]

However, when the degree of K/k is prime to p, it is possible to make the strong statement:

THEOREM I. *If K/k is normal, of degree prime to p, there are 1-units ϵ_1, $\epsilon_2, \cdots, \epsilon_n$ such that every 1-unit ϵ is expressible in the form*

$$\epsilon = \zeta^{\nu}\epsilon_1^{\gamma_1}\epsilon_2^{\gamma_2}\cdots\epsilon_n^{\gamma_n},$$

where ζ is a primitive p^s-th root of maximum order, $\gamma_i \in \Gamma$, which is the group ring of G over the p-adic integers $\mathfrak{o}$, and n is the degree of k/R_p. Also, the expression is unique, so that

$$\zeta^{\nu}\epsilon_1^{\gamma_1}\epsilon_2^{\gamma_2}\cdots\epsilon_n^{\gamma_n} = 1$$

implies $\nu \equiv 0 \bmod p^s$, $\gamma_1 = \gamma_2 = \cdots = \gamma_n = 0$.

Proof. Again we concentrate our attention on the logarithms of 1-units rather than on the 1-units themselves. As before, we have the isomorphism

$$\frac{\mathrm{H}}{(\zeta)} \cong \log \mathrm{H}.$$

It will now be possible to find an additive basis over Γ for the group of logarithms.

Just as before, form the group $\sum_{\sigma, i} \mathfrak{o}\sigma(\mathrm{A}_i) = \Theta$; this can also be written:

$$\Theta = \Gamma\mathrm{A}_1 + \Gamma\mathrm{A}_2 + \cdots + \Gamma\mathrm{A}_n.$$

The expression for elements of Θ is unique, since the $\sigma(\mathrm{A}_i)$ form an independent basis over R_p.

By its construction Θ is a left Γ-modulus and from (2) on page 264 contains $\log \mathrm{H}$ as subgroup of finite index; $\log \mathrm{H}$ is also a Γ-modulus, since $\Gamma \log \mathrm{H} = \log \mathrm{H}$. Now, in the special case when the order of G is prime to p, the group ring Γ is a principal ideal ring (the proof of this fact will be outlined in the next section). It follows that $\log \mathrm{H}$ has an independent basis[8] over Γ of n elements, say $\log \epsilon_1, \log \epsilon_2, \cdots, \log \epsilon_n$. Thus,

[7] The statement of Herbrand's lemma is:

Let $\mathfrak{G}$ be a group, T_1 and T_2 homomorphisms of $\mathfrak{G}$ into itself, such that $T_1T_2\mathfrak{G} = T_2T_1\mathfrak{G} = 1$; let $\mathfrak{G}_1$ be the subgroup for which $T_1\mathfrak{G}_1 = 1$, $\mathfrak{G}_2$ the subgroup for which $T_2\mathfrak{G}_2 = 1$. If, now, $\mathfrak{g}$ is a subgroup of $\mathfrak{G}$, $(\mathfrak{G}:\mathfrak{g})$ finite, such that T_1 and T_2 take $\mathfrak{g}$ also into itself, and $\mathfrak{g}_1$, $\mathfrak{g}_2$ are defined in $\mathfrak{g}$ just as $\mathfrak{G}_1$, $\mathfrak{G}_2$ were in $\mathfrak{G}$, then

$$\frac{(\mathfrak{G}_2:T_1\mathfrak{G})}{(\mathfrak{G}_1:T_2\mathfrak{G})} = \frac{(\mathfrak{g}_2:T_1\mathfrak{g})}{(\mathfrak{g}_1:T_2\mathfrak{g})}.$$

For local class field theory, $\mathfrak{G}$ would be the group of 1-units, $\mathfrak{g}$ the subgroup having a normal basis.

[8] Here the well-known theorem is used that any submodulus of a modulus with a finite basis over a principal ideal ring also has a basis. Generally, it might occur that the basis elements of the submodulus have annihilating ideals other than (0), since the ring (Γ in this case) contains divisors of zero. However, in our case this cannot occur, as the final argument shows.

$$\log \mathrm{H} = \Gamma \log \epsilon_1 + \Gamma \log \epsilon_2 + \cdots + \Gamma \log \epsilon_n = \log \epsilon_1^\Gamma \epsilon_2^\Gamma \cdots \epsilon_n^\Gamma,$$

and hence

$$\mathrm{H} = (\zeta)\epsilon_1^\Gamma \epsilon_2^\Gamma \cdots \epsilon_n^\Gamma;$$

every element ϵ of H is thus in the form

$$\epsilon = \zeta^\nu \epsilon_1^{\gamma_1} \epsilon_2^{\gamma_2} \cdots \epsilon_n^{\gamma_n}; \qquad \gamma_i \in \Gamma.$$

No non-trivial relations can exist of the sort

$$\zeta^\nu \epsilon_1^{\gamma_1} \epsilon_2^{\gamma_2} \cdots \epsilon_n^{\gamma_n} = 1,$$

for this would say that the N elements ϵ_i^σ which generate H over $\mathfrak{o}$ are not independent, whereas the existence of a normal basis of n elements (over Γ) for a subgroup of H implies that there are at least N independent 1-units over $\mathfrak{o}$ in any set of generators for H. This proves the theorem.

In the special case that $k = R_p$ and K/R_p is a normal extension of degree prime to p, this gives the theorem that every 1-unit is uniquely expressible in the form, $\xi = \zeta^\nu \epsilon^\gamma$, where $\gamma = a\sigma + b\tau + \cdots$, with $a, b, \cdots \in \mathfrak{o}$. In terms of the $\mathfrak{P}$-adic logarithm, this states that every logarithm can be uniquely written as

$$\begin{aligned} \log \xi &= a \log \sigma\epsilon + b \log \tau\epsilon + \cdots \\ &= a\sigma \log \epsilon + b\tau \log \epsilon + \cdots . \end{aligned}$$

In other words, the value domain of the $\mathfrak{P}$-adic logarithm in absolutely normal fields is an $\mathfrak{o}$-lattice, generated over $\mathfrak{o}$ by the conjugates of the logarithm of a single 1-unit (or by the logarithm of the conjugates).

In relative normal fields K/k of degree prime to p, the conclusion is almost the same, except that, instead of one basis element, n are needed, where n is the degree k/R_p.

We shall see by simple counterexamples that such a representation is possible only part of the time, indicating that the value domain of the logarithm in general local fields is much more complicated than in the instance treated above.

2. **The group ring** Γ. In the preceding section, the proof of Theorem I depended on the fact that Γ is a principal ideal ring. I outline the proof of the following more inclusive theorem:

Let k be any $\mathfrak{p}$-adic field, $\mathfrak{o}$ its ring of integers, and G any group of order prime to p. Then the group ring $\mathfrak{o}(G)$, generated by G over $\mathfrak{o}$, is a principal ideal ring.

Proof. The proof requires the theory of algebras. The group ring $k(G)$, where G is taken over k, is a semi-simple algebra. By the Wedderburn structure theorems for semi-simple algebras [6], $k(G)$ is the direct sum of simple two-sided ideals, $k(G) = \mathfrak{A}_1 + \mathfrak{A}_2 + \cdots + \mathfrak{A}_r$; each of these simple algebras $\mathfrak{A}$ can be represented as a total matric algebra in a division algebra D over k. Since k is a $\mathfrak{p}$-adic field, its valuation can be continued in a unique way to any finite extension field, whether commutative or not; hence the division algebra D

can be valuated. Thus D has a unique maximal integral domain $\mathfrak{o}_D$, which is a principal ideal ring; if $\mathfrak{P}$ is the prime ideal in D, then $\mathfrak{o}_D$ consists of those elements x such that $|x|_{\mathfrak{P}} \leqq 1$.

One proves easily that a total matric ring in $\mathfrak{o}_D$ is a maximal domain of the total matric algebra in D. Let $\mathfrak{O}$ be such a maximal domain. Since $\mathfrak{o}_D$ is a principal ideal ring, it follows, without much difficulty, that $\mathfrak{O}$ is also a principal ideal ring [3].

There are infinitely many representations for the simple algebra $\mathfrak{A}$ as matric algebra in D, these representations differing by an inner automorphism. The following argument shows that every maximal domain of $\mathfrak{A}$ is of the same type as $\mathfrak{o}$, that is, is the ring of matrices in $\mathfrak{o}_D$ for some representation of $\mathfrak{A}$ as a total matric algebra in D. Let $\mathfrak{O}'$ be an arbitrary maximal domain of $\mathfrak{A}$. Then $\mathfrak{O}\mathfrak{O}'$ is a left $\mathfrak{O}$-ideal, hence $\mathfrak{O}\mathfrak{O}' = \mathfrak{O}\alpha$, since $\mathfrak{O}$ is a principal ideal ring. Then $\mathfrak{O}\mathfrak{O}'$ is a right $\mathfrak{O}'$-ideal, so $\mathfrak{O}\alpha\mathfrak{O}' \subset \mathfrak{O}\alpha$, and consequently $\alpha\mathfrak{O}' \subset \mathfrak{O}\alpha$, therefore $\mathfrak{O}' \subset \alpha^{-1}\mathfrak{O}\alpha$. Since $\mathfrak{O}$ is a domain, so is $\alpha^{-1}\mathfrak{O}\alpha$, and since $\mathfrak{O}'$ is maximal, it follows that $\mathfrak{O}' = \alpha^{-1}\mathfrak{O}\alpha$. Therefore, $\mathfrak{O}'$ is a maximal domain of the same type as $\mathfrak{O}$ in the representation for $\mathfrak{A}$ obtained by transforming with α.

Now, in the group ring $k(G)$, we have that $\mathfrak{o}(G)$ is a maximal domain. This is seen as follows. The order of G is prime to p, hence is a unit in $\mathfrak{o}$; the discriminant of $\mathfrak{o}(G)$, which is a power of the group order, is then also a unit in $\mathfrak{o}$. Consequently, $\mathfrak{o}(G)$ is contained in no larger domain, for otherwise the discriminant of $\mathfrak{o}(G)$ would have to contain a square factor other than a unit. From this it follows that the components of $\mathfrak{o}(G)$ in the simple algebras of $k(G)$ are maximal domains.

The above arguments show directly that $\mathfrak{o}(G)$ is the direct sum of principal ideal rings, from which it is obvious that $\mathfrak{o}(G)$ itself is a principal ideal ring.

3. **Fields without a normal basis for 1-units.** Krasner has proved the following theorem.

In a regular field K/k (that is, not containing a primitive p-th root of unity), the 1-units have a normal basis of n elements if K/k has no higher ramification (n is the degree k/R_p).

I show the converse here.

In any regular field K/k with higher ramification, there exists no normal basis for the 1-*units.*

LEMMA 1. *If the* 1-*units of K/k have a normal basis, then all* 1-*units of k are norms.*

Proof. Let $\epsilon = \epsilon_1^{\gamma_1}\epsilon_2^{\gamma_2}\cdots\epsilon_n^{\gamma_n}$ be a 1-unit of k, the ϵ_i being the normal basis for the 1-units of K. Then ϵ is left invariant by the automorphisms of K/k. But since the representation for ϵ is unique, $\sigma\gamma_i = \tau\gamma_i = \cdots$ (σ, τ, $\cdots$ being all the automorphisms of K/k). This means that $\gamma_i = a_i\,(\sigma + \tau + \cdots)$, or

$$\epsilon = (N\epsilon_1)^{a_1}(N\epsilon_2)^{a_2}\cdots(N\epsilon_n)^{a_n} = N(\epsilon_1^{a_1}\epsilon_2^{a_2}\cdots\epsilon_n^{a_n}),$$

where N stands for "norm".

LEMMA 2. *If K/k is purely ramified of degree p, then the index $(\epsilon:NE) = p$, where ϵ equals all 1-units of k, E equals all 1-units of K.*

Proof. For this we use the theorem[9] that in local cyclic fields, $(u:NU) = e$, where u equals all units of k, U equals all units of K, and e is the ramification index of K/k; in this case, since K/k is purely ramified, $e = p$. Let the number of residue classes mod $\mathfrak{P}$ be p^f; then k contains the $p^f - 1$ roots of unity, which form a complete system of representatives for the multiplicative residue classes mod $\mathfrak{P}$. Hence, if η runs through the $p^f - 1$ roots of unity, then $u = \eta\epsilon$, $U = \eta E$. So,

$$p = (u:NU) = (\eta\epsilon:\eta NE) = (\epsilon:NE),$$

proving the lemma.

To prove the theorem, now let K/k have higher ramification. Suppose K/k did have a normal basis for its 1-units. Since the ramification order is divisible by p, there is a field $\bar{K}$ under K such that K is purely ramified of degree p over $\bar{K}$.[10] We contend that the 1-units of K have a normal basis over $\bar{K}$. Let $\epsilon_1, \cdots, \epsilon_n$ be the assumed basis for the 1-units of K with group ring Γ as exponent domain; Γ is the set of elements $\sum_{\sigma \in G} a_\sigma \sigma$. If $\bar{G}$ is the group of $K/\bar{K}$, let $G = \bar{G}\lambda + \bar{G}\mu + \cdots$ be a division of G into right cosets with respect to $\bar{G}$, λ, μ, $\cdots$ being a complete set of representatives; then if $\bar{\Gamma}$ is the group ring of $\bar{G}$ taken over $\mathfrak{o}$, it follows that

$$\Gamma = \bar{\Gamma}\lambda + \bar{\Gamma}\mu + \cdots .$$

The units of K are given by:

$$\begin{aligned} E &= \epsilon_1^{\Gamma}\epsilon_2^{\Gamma} \cdots \epsilon_n^{\Gamma} \\ &= \epsilon_1^{\bar{\Gamma}\lambda}\epsilon_1^{\bar{\Gamma}\mu} \cdots \epsilon_2^{\bar{\Gamma}\lambda}\epsilon_2^{\bar{\Gamma}\mu} \cdots \epsilon_n^{\bar{\Gamma}\lambda}\epsilon_n^{\bar{\Gamma}\mu} \cdots \\ &= (\epsilon_1^{\lambda})^{\bar{\Gamma}}(\epsilon_1^{\mu})^{\bar{\Gamma}} \cdots (\epsilon_2^{\lambda})^{\bar{\Gamma}}(\epsilon_2^{\mu})^{\bar{\Gamma}} \cdots (\epsilon_n^{\lambda})^{\bar{\Gamma}}(\epsilon_n^{\mu})^{\bar{\Gamma}} \cdots . \end{aligned}$$

Furthermore, the ϵ_i^{λ}, ϵ_i^{μ}, $\cdots$ are independent over $\bar{\Gamma}$, as one sees readily by reversing the above steps. Hence, the elements ϵ_i^{λ}, ϵ_i^{μ}, $\cdots$ are a normal basis for the 1-units of $K/\bar{K}$. From Lemma 1 (taking $\bar{K}$ instead of k), it follows that all 1-units of $\bar{K}$ are norms, in contradiction to Lemma 2, for, according to the latter, a purely ramified extension of degree p (as in $K/\bar{K}$) sends its norms of 1-units into a proper subgroup (of index p) in the group of 1-units of $\bar{K}$. This proves the theorem.

4. **Counterexamples.** From the preceding sections we have seen that any

[9] For a proof of this important theorem of local class field theory, see [1]. It is this theorem for which the subgroup of section 1 can be used (see footnote 7).

[10] Proof of this fact requires the theory of the inertial group (the group T leaving fixed the largest subfield unramified over k). The order of T is e; since p divides e, then by Sylow's theorem, there is a subgroup of order p, which therefore leaves fixed a subfield $\bar{K}$ such that $K/\bar{K}$ is purely ramified of degree p.

regular field contains a normal basis for 1-units if and only if the field is without higher ramification. If the field is irregular, then a normal basis exists if its relative degree is not divisible by p. That a simple more general theorem than the latter is not to be expected is seen for the examples of the following two fields:

(1) $R_2(2^{\frac{1}{2}})$. This field is of degree 2, and ramified over R_2. The field is irregular, containing the second roots of unity. The 1-units of this field do not have a normal basis; however, this is not surprising since the field has higher ramification.

(2) $R_2((-3)^{\frac{1}{2}})$. This field is of degree 2, unramified over R_2, and also has no normal basis. This shows that ramification is not the sole conditioning factor for the existence of a normal basis for 1-units.

It will be necessary first to express the 1-units of these fields in a simple way.

(1) Consider $R_2(2^{\frac{1}{2}})/R_2$; its prime ideal is $(2^{\frac{1}{2}})$. Every 1-unit of the field, as will be shown, can be expressed as a power product, with 2-adic integers as exponents, of the 1-units $1 + 2^{\frac{1}{2}}$, 5 (and possibly a factor -1).

Any 1-unit can be written in the form

$$1 + a_1 2^{\frac{1}{2}} + a_2(2^{\frac{1}{2}})^2 + \cdots,$$

where the a_i are units of the field or 0; a 1-unit in which a_k is the first non-zero coefficient is said to be *of order* k.

In this field, since there are only two residue classes of integers mod $(2^{\frac{1}{2}})$, any 1-unit of order k is a representative for the entire set of 1-units of order k. Let the set $\eta_k(k = 1, 2, \cdots)$ be such representatives for the 1-units of every order. Then every 1-unit is uniquely expressible as a product of the η_k. For, let ϵ be any 1-unit, say of order r; then $\epsilon \equiv \eta_r \bmod (2^{\frac{1}{2}})^{r+1}$; hence $\epsilon = \eta_r\epsilon_s$, where ϵ_s is a 1-unit of order s, $s \geqq r + 1$; $\epsilon_s \equiv \eta_s \bmod (2^{\frac{1}{2}})^{s+1}$, so $\epsilon = \eta_r\eta_s\epsilon_t$, $t \geqq s + 1$; proceeding in this way, we obtain $\epsilon = \eta_r\eta_s\eta_t \cdots$, which is unique. Now we see that $1 + 2^{\frac{1}{2}}$ and 5 generate 1-units of every order: $1 + 2^{\frac{1}{2}}$ is of order 1; $(1 + 2^{\frac{1}{2}})^2$ is of order 2; $-(1 + 2^{\frac{1}{2}})^2 = 1 - (2^{\frac{1}{2}})^3 - (2^{\frac{1}{2}})^4 = \alpha$ is of order 3; the order of α^{2^r} is easily seen to be $2r + 3$ for all r; hence the powers of $1 + 2^{\frac{1}{2}}$ generate 1-units of every odd order. $5 = 1 + (2^{\frac{1}{2}})^4$ is of order 4; the order of 5^{2^r} is $2r + 4$; consequently, 5 generates 1-units of all even orders $\geqq 4$. Hence, between $1 + 2^{\frac{1}{2}}$ and 5 (with possibly a factor of -1), 1-units of every order are generated. Thus, every 1-unit can be written:[11]

$$\epsilon = \pm(1 + 2^{\frac{1}{2}})^a 5^b,$$

where a and b are 2-adic integers. Such an expression is unique, for any relation of the sort $\pm(1 + 2^{\frac{1}{2}})^a 5^b = 1$ would imply that $1 + 2^{\frac{1}{2}}$ and 5 can generate 1-units of the same order, which is clearly impossible from the above.

Now assume that the 1-units of $R_2(2^{\frac{1}{2}})/R_2$ could be generated by a normal basis, and let the generating 1-unit be $\pm(1 + 2^{\frac{1}{2}})^a 5^b$; a and b are 2-adic integers. Then,

[11] This is a special instance of a general theorem by Hensel which expresses the 1-units of a $\mathfrak{P}$-adic field by means of a finite basis over the p-adic integers; see [4].

since every 1-unit is assumed to be uniquely expressible in the form $\pm\ [(1 + 2^{\frac{1}{2}})^a 5^b]^{\alpha+\beta\sigma}$, where σ is the automorphism of $R_2(2^{\frac{1}{2}})/R_2$, $\sigma^2 = 1$, we also have

$$\pm\ [(1 + 2^{\frac{1}{2}})^a 5^b]^{\alpha+\beta\sigma} = 5$$

for some 2-adic integers α, β. Since

$$5 = 5^\sigma = \pm\ [(1 + 2^{\frac{1}{2}})^a 5^b]^{\alpha\sigma+\beta}$$

and the representation must be unique, then $\alpha\sigma + \beta = \alpha + \beta\sigma$ or $\alpha = \beta$. Hence,

$$5 = \pm[(1 + 2^{\frac{1}{2}})^a 5^b]^{\alpha(1+\sigma)}$$

or

$$5 = \pm(-1)^a 5^{2b}.$$

$$\therefore\ \pm 5^{2b\alpha-1} = 1.$$

It follows that $2b\alpha - 1 = 0$, which is impossible with b and α 2-adic integers. This disproves the possibility of a normal basis for the 1-units of $R_2(2^{\frac{1}{2}})/R_2$.

(2) In $R_2((-3)^{\frac{1}{2}})/R_2$, the prime ideal is (2) since the field is unramified. There are four classes of incongruent integers mod (2), which are represented by 0 and the cube roots of unity, 1, ω, ω^2. Now any three incongruent 1-units of order k can be taken as representatives for all 1-units of order k, and, as in the preceding example, every 1-unit is uniquely expressible as a product of these representatives. We shall see first that $1 + 2\omega = (-3)^{\frac{1}{2}}$ and $1 + 4\omega$ generate a complete set of representatives for 1-units of every order and hence generate all 1-units. $1 + 2\omega$ and $-(1 + 2\omega)^2 = 1 + 2$ are of order 1; $(1 + 2)\cdot(1 + 2\omega) = 1 + (1 + \omega)2 + 4\omega$ is also of order 1; these three are all incongruent mod (2). $(1 + 2\omega)^2 = 1 - 4$ and $1 + 4$ are of order 2; $(1 - 4)(1 + 4\omega) = 1 + (\omega - 1)4 - 16\omega$ is a third 1-unit of order 2, incongruent to the others, hence those three are representatives for the 1-units of order 2. One sees readily, as in example (1), that $(1 + 4\omega)^{2^r}$ and $(1 - 4)^{2^r}$ are 1-units of order $r + 2$ for every r; also $(1 + 4\omega)^{2^r}(1 - 4)^{2^r}$ is of order $r + 2$ and these three 1-units being incongruent mod (2) represent all 1-units of order $r + 2$. Thus $1 + 2\omega = (-3)^{\frac{1}{2}}$ and $1 + 4\omega$ generate representatives for 1-units of every order; consequently every 1-unit of $R_2((-3)^{\frac{1}{2}})$ can be expressed in the form[12] $\epsilon = \pm\ ((-3)^{\frac{1}{2}})^a(1 + 4\omega)^b$; a, b are 2-adic integers. One sees readily that this representation is unique.

Suppose the 1-units of $R_2((-3)^{\frac{1}{2}})/R_2$ have a normal basis, the generating element being $\pm\ ((-3)^{\frac{1}{2}})^a(1 + 4\omega)^b$. Then, for some 2-adic integers α, β,

$$\pm\ [((-3)^{\frac{1}{2}})^a(1 + 4\omega)^b]^{\alpha+\beta\sigma} = ((-3)^{\frac{1}{2}}).$$

Now,

$$((-3)^{\frac{1}{2}})^\sigma = -((-3)^{\frac{1}{2}}),\ (1 + 4\omega)^\sigma = 1 + 4\omega^2 = 13(1 + 4\omega)^{-1},$$

so

$$\pm\ ((-3)^{\frac{1}{2}})^{a\alpha+a\beta-1}\ (1 - 4\omega)^{b\alpha-b\beta}13^{b\beta} = 1.$$

[12] See footnote 11.

In R_2, $13 = 1 + 2^2 + 2^3$ can be expressed as $13 = (-3)^c$, where c is a 2-adic integer, since the 1-units of R_2 can be generated by -3. Hence $13 = ((-3)^{\frac{1}{2}})^{2c}$, and thus

$$\pm ((-3)^{\frac{1}{2}})^{2c+a(\alpha+\beta)-1}(1 + 4\omega)^{b(\alpha-\beta)} = 1.$$

From the independence of the elements, $1 + 4\omega$ and $(-3)^{\frac{1}{2}}$, it follows that $b(\alpha - \beta) = 0$; of the two possibilities, $b = 0$ or $\alpha = \beta$, the latter is clearly the only possible one since $b = 0$ implies that $(-3)^{\frac{1}{2}}$ generates all 1-units. However, $\alpha = \beta$ has as consequence that $0 = 2c + a(\alpha + \beta) - 1 = 2c + 2a\alpha - 1$, which is impossible for c, a, and α, 2-adic integers. Consequently, $R_2((-3)^{\frac{1}{2}})/R_2$ cannot have a normal basis for its 1-units despite the fact it is unramified.

BIBLIOGRAPHY

1. C. CHEVALLEY, *La théorie du corps de classes*, Annals of Mathematics, 2nd series, vol. 41(1940), pp. 394–418.

2. H. HASSE, *Bericht über neuere Untersuchungen und Probleme aus der Theorie der algebraischen Zahlkörper*, part II (Reziprozitätsgesetz), Jahresbericht der deutschen Mathematiker-Vereinigung, vol. 36(1927), pp. 233–311.

3. H. HASSE, *Über $\mathfrak{P}$-adische Schiefkörper und ihre Bedeutung für die Arithmetik hyperkomplexer Zahlsysteme*, Mathematische Annalen, vol. 104(1930–1931), pp. 495–534.

4. K. HENSEL, *Die multiplikative Darstellung der algebraischen Zahlen für der Bereich eines beliebigen Primteilers*, Journal für die reine und angewandte Mathematik, vol. 146(1916), pp. 189–215.

5. M. KRASNER, *Sur la représentation exponentielle dans les corps relativement galoisiens de nombres $\mathfrak{P}$-adiques*, Acta Arithmetica, vol. 3(1939), pp. 133–173.

6. B. L. VAN DER WAERDEN, *Moderne Algebra*, vol. II, Berlin, 1930–1931.

INDIANA UNIVERSITY.

AN EXPLICIT FORMULA FOR THE SOLUTION OF THE ULTRAHYPERBOLIC EQUATION IN FOUR INDEPENDENT VARIABLES

By Glynn Owens

1. **Introduction.** Equations of the form

$$\sum_{i=1}^{n} \frac{\partial^2 u}{\partial x_i^2} - \sum_{i=1}^{n} \frac{\partial^2 u}{\partial y_i^2} = 0 \qquad (n \geqq 2, m \geqq 2) \tag{1.1}$$

are referred to as "ultrahyperbolic partial differential equations". In 1901, G. Hamel [2][1] investigating the problem of finding all geometries in which the straight lines are the shortest ones considered an equation equivalent to (1.1) ($n = m = 2$); he established the existence of a function which satisfied the equation in the neighborhood of the intersection of two characteristic hyperplanes and which assumed prescribed initial values on these planes, but the initial values are restricted to be analytic in two of their three arguments. It has only been in recent years that properties of the solutions of (1.1) have been discovered that are not restricted by analyticity.

In 1932, L. Ásgeirsson [1] discovered a mean-value theorem which applies to any twice continuously differentiable solution of (1.1) ($n = m$). By the use of this mean-value theorem, it has been shown that on a non-characteristic hyperplane the values for any solution of (1.1) cannot be arbitrarily assigned so as to furnish a solution.[2]

In 1938, F. John [3], using Ásgeirsson's theorem, was able to determine the most general solution, u, of (1.1) ($n = m = 2$) existing in all space; interpreting the independent variables as suitable functions of the Plücker coordinates of a line in 3-dimensional space, u becomes a function of lines and Ásgeirsson's theorem takes the following form. If the line function u is a solution of (1.1) ($n = m = 2$), then for every hyperboloid H of revolution and of one sheet, the mean values of u for the two families of generating lines of H are equal. Calling a function of the lines of 3-dimensional space with the above property harmonic, John demonstrates that every harmonic line function which is twice continuously differentiable is equivalent to a solution of (1.1) ($n = m = 2$). It is also shown that the line integrals of a sufficiently regular point function form a harmonic line function and that a sufficiently regular harmonic line function is representable as line integrals of a point function.

The present paper considers the equation (1.1) ($n = m = 2$). The variables of this equation are regarded as the coordinates of a 4-dimensional point that varies in a domain defined by an initial hypersurface and the characteristic

Received August 26, 1941; presented to the American Mathematical Society, November 22, 1941.

[1] The numbers in brackets refer to the bibliography.

[2] Ásgeirsson's results and the applications of his mean-value theorem are to be found in the book *Methoden der Mathematischen Physik*, vol. II, by R. Courant and D. Hilbert.

cone. Adapting H. Lewy's [4] generalization of the "Riemann integration method" and applying it to the above domain, assuming the existence of a sufficiently regular solution in this domain, an integral formula is derived that gives the value of the solution at the vertex of the cone. The value at the vertex depends upon the values that the solution and a certain finite number of its derivatives assume on that part of the initial surface that is cut out by the cone. §2 contains the derivation of this formula which involves certain "Riemann functions" that are determined in §3.

Under the above assumptions with regard to the solution of (1.1) ($n = m = 2$), it is shown in §4 that there must necessarily exist two equalities holding for the solution and its normal derivative on the initial hypersurface. That is, at a point on the initial hypersurface the value of the solution and of its normal derivative is dependent upon their values in a certain neighborhood on the initial surface. These results are explicitly carried out in the case of a hyperplane for which the two equalities take the form of two simultaneous integro-differential equations; by the means of an example it is shown that these two necessary conditions are not identically satisfied.

To Professor Hans Lewy under whose guidance this thesis was written, I express my appreciation and thanks.

2. **Generalization of Riemann's method and the explicit formula.** Let the linear differential operator $L[u]$ be defined by the following identity.

$$L[u] \equiv u_{x_1x_1} + u_{x_2x_2} - u_{x_3x_3} - u_{x_4x_4} .$$

Then Green's formula as applied to the expression $L[u]$ may be written as follows.

$$\iiiint_G (vL[u] - uL[v])\, d\tau = \iiint_O \left(v \frac{\partial u}{\partial s} - u \frac{\partial v}{\partial s} \right) do, \tag{2.1}$$

where $d\tau$ is the volume element of the domain G and where do is the surface element of the boundary O of G. The operator $\dfrac{\partial}{\partial s}$ is the well-known transversal derivative[3] and is defined by the expression

$$\frac{\partial}{\partial s} \equiv \frac{\partial x_1}{\partial \nu}\frac{\partial}{\partial x_1} + \frac{\partial x_2}{\partial \nu}\frac{\partial}{\partial x_2} - \frac{\partial x_3}{\partial \nu}\frac{\partial}{\partial x_3} - \frac{\partial x_4}{\partial \nu}\frac{\partial}{\partial x_4}, \tag{2.2}$$

where $\dfrac{\partial x_i}{\partial \nu}$ are the direction cosines of the exteriorly directed normal to the surface O.

If $L[u]$ is equated to zero, we obtain that

$$u_{x_1x_1} + u_{x_2x_2} - u_{x_3x_3} - u_{x_4x_4} = 0, \tag{2.3}$$

that is, the ultrahyperbolic partial differential equation in four independent variables. In order to proceed with the discussion of this equation, we shall

[3] The transversal derivative is also referred to as the conormal derivative.

make use of the characteristic cone of (2.3). The characteristic cone C of (2.3) that has the origin of coordinates P for its vertex is defined as follows.

$$x_1^2 + x_2^2 - x_3^2 - x_4^2 = 0. \tag{2.4}$$

In Green's formula (2.1), let us replace u by $\dfrac{\partial^2 u}{\partial x_i\, \partial x_k}$, for abbreviation $u_{x_i x_k}$, and v by ω_{ik}; if a summation over $1 \leqq i \leqq k \leqq 4$ is now made, we obtain the result

$$\begin{aligned}\iiiint_G \Big\{ \sum_{1 \leqq i \leqq k \leqq 4} [\omega_{ik} L[u_{x_i x_k}] - u_{x_i x_k} L[\omega_{ik}]] \Big\}\, d\tau \\ = \iiint_O \left\{ \sum_{1 \leqq i \leqq k \leqq 4} \left[\omega_{ik} \frac{\partial u_{x_i x_k}}{\partial s} - u_{x_i x_k} \frac{\partial \omega_{ik}}{\partial s} \right] \right\} do.\end{aligned} \tag{2.5}$$

ω_{ik} are functions which will be stipulated presently and G is a bounded, simply-connected, open set that is defined by an initial hypersurface Γ and the characteristic cone C. We assume of course that the vertex P of C belongs to the boundary O of G and that it does not lie on Γ. For the sake of explicitness, it is assumed that G lies on a particular side of the cone C, namely, the side for which the following inequality holds for all points of G.

$$x_1^2 + x_2^2 - x_3^2 - x_4^2 > 0. \tag{2.6}$$

On the characteristic cone C, the result of operating twice with the transversal derivative, which shall be denoted by $\dfrac{\partial^2}{\partial s^2}$, is the following.

$$\frac{\partial^2}{\partial s^2} = \sum_{i=1}^4 \left(\frac{\partial x_i}{\partial \nu} \right)^2 \frac{\partial^2}{\partial x_i^2} + 2 \sum_{i<k=1}^4 \delta_i \delta_k \frac{\partial x_i}{\partial \nu} \frac{\partial x_k}{\partial \nu} \frac{\partial^2}{\partial x_i\, \partial x_k},$$

where $\dfrac{\partial x_i}{\partial \nu}$ are the direction cosines of the exteriorly directed normal to C and where $\delta_1 = \delta_2 = 1$, $\delta_3 = \delta_4 = -1$. The functions ω_{ik} are now required to satisfy the following conditions on C.

$$\omega_{ik} = \begin{cases} \left(\dfrac{\partial x_i}{\partial \nu} \right)^2 & \text{for } i = k \\ 2\delta_i \delta_k \dfrac{\partial x_i}{\partial \nu} \dfrac{\partial x_k}{\partial \nu} & \text{for } i \neq k \end{cases} \qquad (1 \leqq i \leqq k \leqq 4). \tag{2.7}$$

That is, on the characteristic cone we have that

$$\frac{\partial^2}{\partial s^2} \equiv \sum_{1 \leqq i \leqq k \leqq 4} \omega_{ik} \frac{\partial^2}{\partial x_i\, \partial x_k}. \tag{2.8}$$

In addition to the conditions (2.7) it is now demanded that the functions ω_{ik} be solutions of the ultrahyperbolic equation (2.3). That is,

(2.9) $$L[\omega_{ik}] = 0.$$

Accordingly, with the understanding that the solution u of (2.3) is a sufficiently regular function, the equality (2.5) may now be written as

(2.10) $$\iiint\limits_{C+\Gamma} \left\{ \sum_{1 \leq i \leq k \leq 4} \left[\omega_{ik} \frac{\partial u_{x_i x_k}}{\partial s} - u_{x_i x_k} \frac{\partial \omega_{ik}}{\partial s} \right] \right\} do = 0,$$

where the boundary O of G has been replaced by $C + \Gamma$. In order to carry out explicitly the integration over the characteristic cone C, we make use of the following equivalent definition of C.

(2.11) $$\begin{aligned} x_1 &= r \cos \theta, & x_3 &= p \cos \psi, \\ x_2 &= r \sin \theta, & x_4 &= p \sin \psi \qquad (0 \leqq r, p < \infty) \end{aligned}$$

with $r = p = 2^{-\frac{1}{2}}s$. It is to be noted that s is the distance from the origin to the point on the cone whose coordinates are (x_1, x_2, x_3, x_4). It is also noticed that for fixed θ and ψ the locus of the equations (2.11), with $r = p$, is a generator of the cone. Now transversal differentiation on C is easily seen to be equivalent to differentiation with respect to s. That is, differentiation with respect to length along a generator of C is equivalent to the operation $\dfrac{\partial}{\partial s}$. The element of area on C is

$$do = \tfrac{1}{2}s^2 ds\, d\theta\, d\psi,$$

and since by (2.7) $\dfrac{\partial \omega_{ik}}{\partial s} = 0$ on C, the integral over C in (2.10) is

$$\iiint\limits_{C} \left\{ \frac{\partial}{\partial s} \left[\sum_{1 \leq i \leq k \leq 4} \omega_{ik} u_{x_i x_k} \right] \right\} do = \tfrac{1}{2} \iiint\limits_{C} \frac{\partial^3 u}{\partial s^3} s^2\, ds\, d\theta\, d\psi,$$

because by (2.8), on C,

$$\sum_{1 \leq i \leq k \leq 4} \omega_{ik} u_{x_i x_k}$$

is precisely the second derivative of u with respect to s, the distance. Let $\rho = \rho(\theta, \psi)$ represent the distance from the vertex P of C to the point where the generator determined by θ and ψ (see (2.11)) meets the surface Γ; then, by the use of partial integration, the following expression is obtained.

$$\int_0^{\rho} \frac{\partial^3 u}{\partial s^3} s^2\, ds = \rho^2 \frac{\partial^2 u}{\partial s^2} - 2\rho \frac{\partial u}{\partial s} + 2u - 2u(P),$$

where the arguments (x_1, x_2, x_3, x_4) of the first three terms on the right are respectively replaced by $2^{-\frac{1}{2}}\rho \cos \theta$, $2^{-\frac{1}{2}}\rho \sin \theta$, $2^{-\frac{1}{2}}\rho \cos \psi$, $2^{-\frac{1}{2}}\rho \sin \psi$ and where $u(P)$ means the value of u at P. Making use of (2.10), we find that

$$\tfrac{1}{2}\int_0^{2\pi}\int_0^{2\pi}\left(\rho^2\frac{\partial^2 u}{\partial s^2} - 2\rho\frac{\partial u}{\partial s} + 2u\right)d\theta\, d\psi - 4\pi^2 u(P)$$
$$+ \iiint_{\Gamma}\left\{\sum_{1\leqq i\leqq k\leqq 4}\left[\omega_{ik}\frac{\partial u_{x_i x_k}}{\partial s} - u_{x_i x_k}\frac{\partial \omega_{ik}}{\partial s}\right]\right\} do = 0,$$

or

$$\begin{aligned} 8\pi^2 u(P) = {} & \int_0^{2\pi}\int_0^{2\pi}\left(\rho^2\frac{\partial^2 u}{\partial s^2} - 2\rho\frac{\partial u}{\partial s} + 2u\right)d\theta\, d\psi \\ & + 2\iiint_{\Gamma}\left\{\sum_{1\leqq i\leqq k\leqq 4}\left[\omega_{ik}\frac{\partial u_{x_i x_k}}{\partial s} - u_{x_i x_k}\frac{\partial \omega_{ik}}{\partial s}\right]\right\} do, \end{aligned} \tag{2.12}$$

where in the first integral of (2.12) $\dfrac{\partial}{\partial s}$ refers to transversal differentiation on C and in the second integral $\dfrac{\partial}{\partial s}$ refers to transversal differentiation on the surface Γ.

3. **The Riemann functions.** In the interior of the domain G of §2 the functions ω_{ik} satisfy the equation $L[\omega_{ik}] = 0$, that is, the ultrahyperbolic equation, and on the characteristic cone the following characteristic initial values:

$$\omega_{ik} = \begin{Bmatrix} \left(\dfrac{\partial x_i}{\partial \nu}\right)^2 & \text{for } i = k \\ 2\delta_i\,\delta_k\,\dfrac{\partial x_i}{\partial \nu}\,\dfrac{\partial x_k}{\partial \nu} & \text{for } i \neq k \end{Bmatrix} \qquad (1 \leqq i \leqq k \leqq 4). \tag{3.1}$$

See (2.7) and (2.9). Let us now make the following transformation of coordinates:

$$x_1 = r\cos\theta, \qquad x_3 = p\cos\psi,$$
$$x_2 = r\sin\theta, \qquad x_4 = p\sin\psi \qquad (0 \leqq r,\ p < \infty).$$

The equation (2.3) is transformed into

$$u_{rr} + \frac{u_r}{r} + \frac{u_{\theta\theta}}{r^2} - u_{pp} - \frac{u_p}{p} - \frac{u_{\psi\psi}}{p^2} = 0. \tag{3.2}$$

The boundary conditions (3.1) become:

$$\begin{aligned} &\omega_{11} = \frac{\cos^2\theta}{2}, \quad \omega_{12} = \sin\theta\cos\theta, \quad \omega_{13} = \cos\theta\cos\psi, \quad \omega_{14} = \cos\theta\sin\psi, \\ &\omega_{22} = \frac{\sin^2\theta}{2}, \quad \omega_{23} = \sin\theta\cos\psi, \quad \omega_{24} = \sin\theta\sin\psi, \\ &\omega_{33} = \frac{\cos^2\psi}{2}, \quad \omega_{34} = \sin\psi\cos\psi, \quad \omega_{44} = \frac{\sin^2\psi}{2}. \end{aligned} \tag{3.3}$$

Solutions of (3.2) which assume the values (3.3) and which are regular in the domain (2.6) are required. The domain (2.6) is equivalent to the domain $r > p \geqq 0$. The boundary C of (2.6) is equivalent to $r = p$. To find these functions it suffices to try one or the other of the two following forms for ω_{ik}.

$$\omega_{ik} = \text{constant} + f(\theta, \psi) g\left(\frac{p^2}{r^2}\right)$$

or

$$\omega_{ik} = \text{constant} + f(\theta, \psi) g\left(\frac{p^2}{r^2}, r^2 - p^2\right).$$

The resulting Riemann functions ω_{ik} are given in the following table:

$$\omega_{11} = \frac{1}{4} + \frac{\cos 2\theta}{4}\left[1 + \left(1 - \frac{p^2}{r^2}\right)\log(r^2 - p^2)\right];$$

$$\omega_{12} = \frac{\sin 2\theta}{2}\left[1 + \left(1 - \frac{p^2}{r^2}\right)\log(r^2 - p^2)\right];$$

$$\omega_{13} = \frac{p}{r}\cos\theta\cos\psi;$$

$$\omega_{14} = \frac{p}{r}\cos\theta\sin\psi;$$

$$\omega_{22} = \frac{1}{4} - \frac{\cos 2\theta}{4}\left[1 + \left(1 - \frac{p^2}{r^2}\right)\log(r^2 - p^2)\right];$$

$$\omega_{23} = \frac{p}{r}\sin\theta\cos\psi;$$

$$\omega_{24} = \frac{p}{r}\sin\theta\sin\psi;$$

$$\omega_{33} = \frac{1}{4} + \frac{\cos 2\psi}{4}\left[1 + \left(\frac{r^2}{p^2} - 1\right)\log\left(1 - \frac{p^2}{r^2}\right)\right];$$

$$\omega_{34} = \frac{\sin 2\psi}{2}\left[1 + \left(\frac{r^2}{p^2} - 1\right)\log\left(1 - \frac{p^2}{r^2}\right)\right];$$

$$\omega_{44} = \frac{1}{4} - \frac{\cos 2\psi}{4}\left[1 + \left(\frac{r^2}{p^2} - 1\right)\log\left(1 - \frac{p^2}{r^2}\right)\right].$$

It is to be noticed that the trigonometric terms in ω_{33}, ω_{34}, and ω_{44} are undefined for $p = 0$, but ω_{33}, ω_{34} and ω_{44} are defined. This follows from the fact that the common term multiplying the trigonometric factors has the factor p^2. One observes that the derivatives of certain of the ω_{ik} with respect to the variables x_i are infinite on the cone C. Hence the following remarks are necessary to justify the validity of the use of Green's formula in §2. Consider the domain defined by B, that is,

$$x_1^2 + x_2^2 - x_3^2 - x_4^2 \geqq a^2$$

and the surface Γ of §2. As $a \to 0$, this approximating domain tends to that defined by C and Γ. Green's formula is valid in this domain which lies interior to our original domain. On the boundary B^* of B, the transversal derivatives of ω_{33}, ω_{34} and ω_{44} are zero by virtue of the homogeneity of these functions. The transversal derivatives on Γ of the six ω_{ik} mentioned above are logarithmically infinite at the intersection of C and Γ; hence the surface integrals of

$$\left(\omega_{ik} \frac{\partial u_{x_i x_k}}{\partial s} - u_{x_i x_k} \frac{\partial \omega_{ik}}{\partial s}\right)$$

over the boundary of the approximating domain that lies on Γ tends to a limit as $a \to 0$. Although the transversal derivatives of ω_{11}, ω_{12} and ω_{22} are not zero on B^*, nevertheless, it is an easy consideration to show that as $a \to 0$

$$\lim_{a \to 0} \iiint_{B^*} \frac{\partial \omega_{ik}}{\partial s} u_{x_i x_k}\, do = 0.$$

Consequently, the manner in which Green's formula has been used is justified.

4. **Necessary conditions for the initial values.** It has been shown by means of Ásgeirsson's mean-value theorem that an arbitrary assignment of initial values on a non-characteristic hyperplane does not lead to a solution of the ultrahyperbolic equation. I propose to investigate further this question of suitable initial values. Necessary conditions will be derived for the following problem. What initial values are permissible in order to ensure a solution of the equation (2.3)?

Let us commence by considering Green's formula (2.5) as used in §2, namely,

$$\begin{aligned} &\iiiint_G \left\{ \sum_{1 \leqq i \leqq k \leqq 4} [\omega_{ik} L[u_{x_i x_k}] - u_{x_i x_k} L[\omega_{ik}]] \right\} d\tau \\ &\qquad = \iiint_G \left\{ \sum_{1 \leqq i \leqq k \leqq 4} \left[\omega_{ik} \frac{\partial u_{x_i x_k}}{\partial s} - u_{x_i x_k} \frac{\partial \omega_{ik}}{\partial s}\right] \right\} do, \end{aligned} \tag{4.1}$$

where the notation has the same meaning as in §2 except that now the vertex, the origin of coordinates, P of C lies on the surface Γ. By the results of §2, it is permissible to write (4.1) as

$$\iiint_\Gamma \left\{ \sum_{1 \leqq i \leqq k \leqq 4} \left[\omega_{ik} \frac{\partial u_{x_i x_k}}{\partial s} - u_{x_i x_k} \frac{\partial \omega_{ik}}{\partial s}\right] \right\} do + \frac{1}{2} \iiint_C \frac{\partial^3 u}{\partial s^3} s^2\, ds\, d\theta\, d\psi = 0.$$

By the use of partial integration as in §2, we find that

$$\frac{1}{2} \iiint_C \frac{\partial^3 u}{\partial s^3} s^2\, ds\, d\theta\, d\psi = \frac{1}{2} \iint \left(\rho^2 \frac{\partial^2 u}{\partial s^2} - 2\rho \frac{\partial u}{\partial s} + 2u\right) d\theta\, d\psi - u(P) \iint d\theta\, d\psi,$$

and accordingly our first equality is obtained. Namely,

$$I(\theta, \psi)u(P) = \frac{1}{2}\iint \left(\rho^2 \frac{\partial^2 u}{\partial s^2} - 2\rho \frac{\partial u}{\partial s} + 2u\right) d\theta\, d\psi + \iiint_{\Gamma} \left\{ \sum_{1 \leq i \leq k \leq 4} \left[\omega_{ik} \frac{\partial u_{x_i x_k}}{\partial s} - u_{x_i x_k} \frac{\partial \omega_{ik}}{\partial s} \right] \right\} do, \tag{4.2}$$

where $I(\theta, \psi) = \iint d\theta\, d\psi$ and where this double integral depends only on Γ and P. Now to obtain the value of $u_{x_i}(P)$, we need merely to substitute u_{x_i} for u in equation (4.2), and for convenience let $v \equiv u_{x_i}(P)$. One finds that

$$I(\theta, \psi)u_{x_i}(P) = \frac{1}{2}\iint \left(\rho^2 \frac{\partial^2 v}{\partial s^2} - 2\rho \frac{\partial v}{\partial s} + 2v\right) d\theta\, d\psi + \iiint_{\Gamma} \left\{ \sum_{1 \leq i \leq k \leq 4} \left[\omega_{ik} \frac{\partial v_{x_i x_k}}{\partial s} - v_{x_i x_k} \frac{\partial \omega_{ik}}{\partial s} \right] \right\} do. \tag{4.3}$$

If $\dfrac{\partial x_i}{\partial \nu}$ denote the direction cosines of the normal to Γ, then the normal derivative $\dfrac{\partial u}{\partial \nu}$ on Γ is given by

$$\frac{\partial u}{\partial \nu} = \sum \frac{\partial x_i}{\partial \nu} \frac{\partial u}{\partial x_i}. \tag{4.4}$$

Equations (4.3) and (4.4) determine an equality which the normal derivative of u must satisfy on Γ.

A definite case is now to be treated. A domain $G(R)$ that is defined by a hypersphere of radius R, the characteristic cone with vertex at P, and the hyperplane $x_1 = 0$ is to be considered. If the transformation of coordinates

$$\begin{aligned} x_1 &= r \cos \theta, & x_3 &= p \cos \psi, \\ x_2 &= r \sin \theta, & x_4 &= p \sin \psi \qquad (0 \leqq r, p < \infty) \end{aligned} \tag{4.5}$$

is made, then the analytical definition of $G(R)$ is

$$\begin{gathered} r^2 + p^2 < R^2, \\ r^2 - p^2 > 0, \\ \cos \theta > 0. \end{gathered} \tag{4.6}$$

If the domain $G(R)$ is taken as the region G of (4.1), then it is found that

$$\iiint_{S+\bar{P}} \left\{ \sum_{1 \leq i \leq k \leq 4} \left[\omega_{ik} \frac{\partial u_{x_i x_k}}{\partial s} - u_{x_i x_k} \frac{\partial \omega_{ik}}{\partial s} \right] \right\} do + \frac{1}{2} \iiint_{C} \frac{\partial^3 u}{\partial s^3} s^2\, ds\, d\theta\, d\psi = 0. \tag{4.7}$$

S, C and $\bar{P}$ mean, respectively, that part of the boundary of $G(R)$ that lies on the boundary of the sphere, the cone and the plane. The arguments of u in the last integral as in §2 are (4.5) with $r = p = 2^{-\frac{1}{2}}s$. For fixed θ and ψ and with $r = p = 2^{-\frac{1}{2}}s$, (4.5) defines a generator of the characteristic cone passing through P. This generator meets the surface S at a distance $s = R$ from P. Accordingly, it is permissible to write

$$\int_0^R \frac{\partial^3 u}{\partial s^3} s^2\, ds = R^2 \frac{\partial^2 u}{\partial s^2} - 2R \frac{\partial u}{\partial s} + 2u(R) - 2u(P).$$

Since $\cos\theta > 0$, $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$ and (4.7) becomes

$$\tag{4.8} \begin{aligned} &\frac{1}{2}\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} d\theta \int_0^{2\pi} \left(R^2 \frac{\partial^2 u}{\partial s^2} - 2R\frac{\partial u}{\partial s} + 2u\right) d\psi - 2\pi^2 u(P) \\ &\qquad + \iiint\limits_{S+\bar{P}} \left\{ \sum_{1 \le i \le k \le 4} \left[\omega_{ik} \frac{\partial u_{x_i x_k}}{\partial s} - u_{x_i x_k} \frac{\partial \omega_{ik}}{\partial s} \right] \right\} do = 0. \end{aligned}$$

Assume now that u and its derivatives vanish sufficiently rapidly as R tends to infinity. This assures the vanishing of all integrals of (4.8) that are taken on S. Hence,

$$\tag{4.9} 2\pi^2 u(P) = \iiint\limits_A \left\{ \sum_{1 \le i \le k \le 4} \left[u_{x_i x_k} \frac{\partial \omega_{ik}}{\partial s} - \omega_{ik} \frac{\partial u_{x_i x_k}}{\partial s} \right] \right\} dx_2\, dx_3\, dx_4,$$

where A is the 3-dimensional domain $x_2^2 - p^2 > 0$. Placing $v \equiv u_{x_1}$, we find that

$$\tag{4.10} 2\pi^2 u_{x_1}(P) = \iiint\limits_A \left\{ \sum_{1 \le i \le k \le 4} \left[v_{x_i x_k} \frac{\partial \omega_{ik}}{\partial s} - \omega_{ik} \frac{\partial v_{x_i x_k}}{\partial s} \right] \right\} dx_2\, dx_3\, dx_4.$$

Therefore the initial values satisfy a pair of simultaneous integro-differential equations.

If (4.9) and (4.10) were identically satisfied by all functions u and v which are sufficiently regular and which vanish properly at infinity, then these equations would not be restrictions on the initial values. It suffices to show that (4.9) is not satisfied for arbitrary v, subject to the above conditions. Consider that part, $M(v)$, of the integrand of (4.9) that does not contain u or its derivatives.[4] Namely,

$$\tag{4.11} M(v) = -\left\{ \begin{aligned} &\omega_{11}(v_{x_2x_2} - v_{x_3x_3} - v_{x_4x_4}) + \frac{\partial \omega_{12}}{\partial x_1} v_{x_2} + \frac{\partial \omega_{13}}{\partial x_1} v_{x_3} \\ &\frac{\partial \omega_{14}}{\partial x_1} v_{x_4} - \omega_{22} v_{x_2x_2} - \omega_{23} v_{x_2x_3} - \omega_{24} v_{x_2x_4} \\ &-\omega_{33} v_{x_3x_3} - \omega_{34} v_{x_3x_3} - \omega_{44} v_{x_4x_4} \end{aligned} \right\}.$$

[4] u and v are initial values prescribed on the plane $x_1 = 0$ and the operator $\frac{\partial}{\partial s} = -\frac{\partial}{\partial x_1}$ on $x_1 = 0$.

Since (4.9) is now assumed to hold for arbitrary v, the integral $I(v)$ of $M(v)$ over the interior of the cone A must vanish for arbitrary v. v will now be defined so as to provide a contradiction. Let

$$(4.12)\qquad v = \left\{\begin{array}{ll} (l^2 - x_2^2)^\alpha & \text{for } x_2^2 \leqq l^2 \text{ and } r^2 + p^2 \leqq \rho^2 \\ 0 & \text{for } x_2^2 \geqq l^2 \text{ and } r^2 + p^2 \leqq \rho^2 \\ \multicolumn{2}{l}{\text{Continue } v \text{ outside of the circle } r^2 + p^2 \leqq \rho^2 \text{ in a sufficiently}} \\ \multicolumn{2}{l}{\text{regular manner so that } v \text{ vanishes outside of a circle } r^2 + p^2 =} \\ \multicolumn{2}{l}{\delta^2, \text{ where } \delta > \rho.} \end{array}\right\},$$

where α is a sufficiently large number. It is now to be shown that $I(v)$ does not vanish. Making use of (4.12) one finds that

$$(4.13)\qquad I(v) = \iiint_A \left[(\omega_{22} - \omega_{11})v_{x_2x_2} - \frac{\partial\omega_{12}}{\partial x_1} v_{x_2}\right] dx_2\,dx_3\,dx_4,$$

and that for $x_1 = 0$,

$$(4.14)\qquad \begin{aligned} \omega_{22} - \omega_{11} &= \frac{1}{2}\left[1 + \left(1 - \frac{p^2}{x_2^2}\right)\log\,(x_2^2 - p^2)\right], \\ \frac{\partial\omega_{12}}{\partial x_1} &= \frac{1}{x_2}\left[1 + \left(1 - \frac{p^2}{x_2^2}\right)\log\,(x_2^2 - p^2)\right]. \end{aligned}$$

Substituting the quantities in (4.14) into (4.13), $I(v)$ becomes

$$(4.15)\quad I(v) = \iiint_A \left[1 + \left(1 - \frac{p^2}{x_2^2}\right)\log\,(x_2^2 - p^2)\right]\left[\tfrac{1}{2}v_{x_2x_2} - \frac{1}{x_2}v_{x_2}\right] dx_2\,dx_3\,dx_4.$$

By the use of the definition of v, it is found that

$$\tfrac{1}{2}v_{x_2x_2} - \frac{1}{x_2}v_{x_2} = \alpha(l^2 - x_2^2)^{\alpha-1} + 2\alpha(\alpha - 1)x_2^2(l^2 - x_2^2)^{\alpha-2}$$

and hence

$$\begin{aligned} I(v) = 2 \iiint_{x_2 - p > 0} &\left[1 + \left(1 - \frac{p^2}{x_2^2}\right)\log\,(x_2^2 - p^2)\right] \\ &\cdot[\alpha(l^2 - x_2^2)^{\alpha-1} + 2\alpha(\alpha - 1)x_2^2(l^2 - x_2^2)^{\alpha-1}]p\,dp\,d\psi\,dx_2, \end{aligned}$$

observing that the transformation $x_3 = p\cos\psi$, $x_4 = p\sin\psi$ has been made. To show that $I(v)$ does not vanish it suffices to show that the next integral J is zero at $x_2 = 0$ and is monotone.

$$J = \int_0^{x_2}\left[p + p\left(1 - \frac{p^2}{x_2^2}\right)\log\,(x_2^2 - p^2)\right]dp.$$

Evaluating J we find that

$$J = \tfrac{3}{8}x_2^2 + \tfrac{1}{2}x_2^2 \log x_2 . \tag{4.16}$$

From (4.16) it is obvious that J has the desired properties provided that l (see (4.12)) has been chosen sufficiently small. Consequently, the integral $I(v)$ does not vanish and we have a contradiction.

In conclusion, I would like to say that the methods of this paper have been successfully applied to the ultrahyperbolic equation in five independent variables.

Bibliography

1. L. Ásgeirsson, *Über eine Mittelwertseigenschaft von Lösungen homogener linearer partieller Differentialgleichungen 2. Ordnung mit konstanten Koeffizienten*, Mathematische Annalen, vol. 113(1936), pp. 321–346.
2. G. Hamel, *Über die Geometrieen, in denen die Geraden die Kürzesten sind*, Mathematische Annalen, vol. 57(1903), pp. 231–264.
3. F. John, *The ultrahyperbolic differential equation with four independent variables*, Duke Mathematical Journal, vol. 4(1938), pp. 300–322.
4. H. Lewy, *Verallgemeinerung der Riemannschen Methode auf mehr Dimensionen*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, 1928, pp. 118–123.

University of California, at Berkeley.

GENERALIZED ARITHMETIC

By Garrett Birkhoff

1. **Introduction.** Since the time of Cantor, it has been the fashion to divide all arithmetic at the root into two separate branches: cardinal and ordinal. Each of these branches is supposed to have its peculiar operations of addition, multiplication, and exponentiation. Only as an afterthought are the two branches connected, by a roughly[1] homomorphic correspondence from ordinal arithmetic to cardinal arithmetic, which is ismorphic when restricted to finite ordinals and cardinals.

In the present paper, an entirely different point of view is advanced. Instead of giving finite and transfinite arithmetic a split personality, half ordinal, half cardinal, I believe that one should regard both aspects as fragments of a unified general arithmetic of partially ordered systems.

I should like to stress three arguments in favor of this point of view.

In the first place, what are usually considered as purely cardinal operations extend in a natural way to ordinal numbers and other partially ordered systems, and vice versa. Moreover, when applied to the wider context of general partially ordered systems, the six operations of "generalized arithmetic" are found to have important new applications.[2] The variety and importance of these will stand comparison with the applications of transfinite arithmetic, as that term has been understood heretofore.[3]

In the second place, almost all arithmetical laws which are valid in transfinite arithmetic, as that term is understood now, are equally valid when the operations are applied to the most general partially ordered systems. In fact, the big gap comes between ordinary arithmetic and transfinite arithmetic; much more is lost by admitting infinite numbers as legitimate objects for arithmetic operations than is lost by including partially ordered sets in the middle ground between totally ordered sets (finite ordinals) and totally unordered sets (finite cardinals). Moreover, the slight loss is more than compensated by the availability of *new cross-laws* connecting cardinal with ordinal operations.

Finally, adoption of the broader point of view towards arithmetic developed below fits the traditional transfinite arithmetic into the general framework of

Received October 18, 1941.

[1] Ordinal exponentiation does not quite fit into this statement, and involves special complications. Also, the proof of this connection involves the well-ordering principle (axiom of choice).

[2] Much may be found about the extension of cardinal operations and applications of the extended definitions in [1]. However, the scope of the present program is nowhere suggested in that paper.

[3] The need for the *operations* of transfinite arithmetic was never very great; the need in topology for even transfinite ordinals has now largely disappeared, thanks to the increased use of the more effective and simpler tool of Moore-Smith convergence.

modern algebra. This gives fresh insight into known facts, and suggests new problems and results.

2. **Numbers, subnumbers, and homonumbers.** Let us agree to mean by the word *number* any non-void partially ordered set. That is ([3], Chap. VI, §2, or [2], p. 5), a "number" is a set A of elements $x, y, z, \cdots$, connected by a reflexive, transitive, and anti-symmetric[4] relation $x \geqq y$. Numbers will be denoted by italic capital letters throughout the sequel.

We shall call two numbers A and B *equal* (in symbols, $A = B$) if and only if they represent *isomorphic* partially ordered sets. This relation has the usual reflexive, symmetric, and transitive properties; moreover, it conforms to accepted usage.

The usual meaning of inequality between cardinal numbers and also that between ordinal numbers appear as special cases of the concept of subnumber as now defined.

Definition. A number A will be called a *subnumber* of a number B (in symbols, $A \subset B$) if and only if A is isomorphic to a *subset* of B.

We shall state without proof the evident

Theorem 1. *The relation of being a subnumber is consistent, reflexive and transitive; unity is a subnumber of every number. Formally,*

(1) *if* $B = C$, *then* $A \subset B$ *implies* $A \subset C$ *and* $B \subset D$ *implies* $C \subset D$;

(2) *for all* A, $A \subset A$;

(3) *if* $A \subset B$ *and* $B \subset C$, *then* $A \subset C$;

(4) *for all* A, $1 \subset A$.

The concept of subnumber may be supplemented by the closely related concept of homonumber, suggested by the general ideas of modern algebra. This concept, in the present context, is new.

Definition. A number A will be called a *homonumber* of a number B (in symbols, $A \prec B$), if and only if there is a many-one or one-one correspondence of B onto A which preserves order, so that if a and a' are the images of b and b', respectively, then $b \geqq b'$ in B implies $a \geqq a'$ in A.

Theorem 2. *The relation of being a homonumber is consistent, reflexive and transitive; unity is a homonumber of every number. Formally,*

(5) *if* $B = C$, *then* $A \prec B$ *implies* $A \prec C$ *and* $B \prec D$ *implies* $C \prec D$;

(6) *for all* A, $A \prec A$;

[4] It has been pointed out to the author by J. W. Tukey that most of the results below are independent of this restriction, and indeed the definition of ordinal exponentiation is simplified. However, this further generalization will not be made here since it might be confusing to many.

(7) *if* $A \prec B$ *and* $B \prec C$, *then* $A \prec C$;

(8) *for all* A, $1 \prec A$.

We note that the ordinal ω is not a homonumber of $\omega + 1$, although it is a subnumber of it. For, any homonumber of $\omega + 1$ would have a last element i, whereas ω has none.

We now come face-to-face with the principal properties of cardinal and ordinal numbers, which are lost in the generalized arithmetic developed here. First, the relation $\subset$, which is *anti-symmetric* for cardinals (Bernstein's theorem) and ordinals, is not anti-symmetric[5] in general. Second, it is obviously not true that for all A, B either $A \subset B$ or $B \subset A$, although this is well known to be true for ordinals ([2], Theorem 1.8), and hence (using the well-ordering principle) for cardinals. In summary, the *comparability* property is lost.[6]

3. **Cardinal and ordinal addition.** The usual definitions of addition for cardinal and ordinal numbers generalize in obvious ways to arbitrary partially ordered sets.[7]

DEFINITION. By the *cardinal sum* of A and B (in symbols, $A + B$) is meant the number consisting of all the elements in A or B, where inclusion within A and within B keep their original meaning, while neither $a \geqq b$ nor $a \leqq b$ holds for any $a \in A$, $b \in B$. By the *ordinal sum* $A \oplus B$ of A and B is meant the number consisting of all elements in A and all those in B, where inclusion within A and within B keep their original meaning, while $a > b$ holds for *all* $a \in A$ and $b \in B$.

Thus for finite numbers, we can construct the diagrams [2] of $A + B$ and of $A \oplus B$ as follows. That of $A + B$ is obtained by laying the diagrams of A and B side-by-side, that of $A \oplus B$ by laying that of A above that of B and drawing lines from all minimal elements of A to all maximal elements of B. Thus the graph of a cardinal sum is always disconnected; while if A has a least element o and B a greatest element i, then the graph of $A \oplus B$ has a "node", or line which, if severed, would disconnect the graph. The converses of these also hold.

THEOREM 3. *Cardinal and ordinal addition are both single-valued, isotone, and associative. Cardinal addition is commutative, and homomorphic to ordinal addition. Formally,*

(9) *if* $A = B$ *and* $C = D$, *then* $A + C = B + D$ *and* $A \oplus C = B \oplus D$;

(10) *if* $A \subset B$ *and* $C \subset D$, *then* $A + C \subset B + D$ *and* $A \oplus C \subset B \oplus D$;

(11) *if* $A \prec B$ *and* $C \prec D$, *then* $A + C \prec B + D$ *and* $A \oplus C \prec B \oplus D$;

[5] Let A consist of all fractions $m + 1/(n + 1)$, where m and n are positive integers; let B consist of all those with $m > 1$ or $n = 1$. Then $A \subset B$ and $B \subset A$, yet $A \neq B$.

[6] Slightly more is lost: the fact that the cardinals and ordinals form well-ordered sets under the relation $A \subset B$.

[7] For the usual definitions, cf. [3] or [4]. The generalizations were outlined in [1]; they are obvious enough.

(12) $A + (B + C) = (A + B) + C$ *and* $A \oplus (B \oplus C) = (A \oplus B) \oplus C$;

(13) $A + B = B + A$;

(14) $A \oplus B \prec A + B$.

Proof. The proofs of (9)–(11) follow by just putting together the correspondences between the summands. Formulas (12)–(13) are obvious; the correspondence giving (14) is just the identical correspondence from A to A and B to B.

Remark. It is known that $\omega = 1 \oplus \omega \neq \omega \oplus 1$: ordinal addition of *infinite* ordinals is not commutative. It is also true that ordinal addition of *finite* cardinals is non-commutative.

Finally, we can state some results connecting inclusion with addition, which are true except for the combination of homomorphic inclusion with ordinal addition. We have

(15) if $A \subset B + C$, then $A = D + E$, where $D \subset B$ and $E \subset C$;

(16) if $A \subset B \oplus C$, then $A = D \oplus E$, where $D \subset B$ and $E \subset C$;

(17) if $B + C \prec A$, then $A = D + E$, where $B \prec D$ and $C \prec E$;

(18) $A \subset A + B$, $B \subset A + B$, $A \subset A \oplus B$, and $B \subset A \oplus B$;

(19) $A \prec A + B$, and $B \prec A + B$.

Of these results, (15)–(17) assert the hereditary nature of decomposability, and (15)–(16) are converses of (10)–(11). While (18) would be trivial if one admitted the void set 0 as a number,[8] it would then be a corollary of (10).

4. **Digression: unique decomposition theorems.** It may be shown that decomposition into either cardinal or ordinal summands is essentially unique.

THEOREM 4. *Any two decompositions of a number A into cardinal summands or ordinal summands have a common refinement.*

COROLLARY. *A number has at most one cardinal or ordinal decomposition into indecomposable summands; in the finite case, this always exists.*

Proof. It is easy to show that a partition of a "number" A represents a decomposition of A into cardinal summands if and only if $x \geqq y$ implies that x and y are in the same subdivision of A. But it is easy to show that the product (in the ordinary sense) of any two partitions with this property itself has this property, and is the desired refinement. Similarly, a partition of A represents a decomposition of A into ordinal summands if and only if, given any two pieces A_i and A_j of the partition, either $x > y$ for all $x \in A_i$, $y \in A_j$, or the reverse holds. With a little trouble, one can also show that the product of any two such partitions is also such a partition.

[8] There are two troubles with this. First, 0^0 and ${}^0 0$ would then be ambiguously defined (they could be either 0 or 1), and second, (8) would no longer be true.

We note that in the case of ordinal sums, even the order of summation is prescribed. With cardinal sums, the representation as a sum of indecomposable summands is only unique to within rearrangement of the factors.

5. **Cardinal multiplication.** The usual notion of the product of two cardinal numbers also has an obvious and known (cf. [1] or [2]) generalization to arbitrary partially ordered sets or, in our terminology, to arbitrary "numbers".

DEFINITION. By the *cardinal product* of two given numbers A and B (written AB) is meant the set of all couples (a, b) $(a \in A, b \in B)$, where $(a, b) \geqq (a', b')$ if and only if $a \geqq a'$ in A and $b \geqq b'$ in B.

THEOREM 5. *Cardinal multiplication is single-valued, isotone, commutative, and associative; it admits* 1 *as an identity, is distributive on cardinal sums, and semi-distributive on ordinal sums. Formally,*

(20) $A = B$ *implies* $AC = BC$ *and* $CA = CB$;

(21) $A \subset B$ *implies* $AC \subset BC$ *and* $CA \subset CB$;

(22) $A \prec B$ *implies* $AC \prec BC$ *and* $CA \prec CB$;

(23) $AB = BA$ *for all* A, B;

(24) $A(BC) = (AB)C$ *for all* A, B, C;

(25) $1A = A1 = A$ *for all* A;

(26) $A(B + C) = AB + AC$ *and* $(A + B)C = AC + BC$;

(27) $A(B \oplus C) \succ AB \oplus AC$ *and* $(A \oplus B)C \succ AC \oplus BC$.

The verification of each of these laws individually, except (27), is a straightforward and elementary exercise which involves only substituting in definitions and using obvious correspondences (cf. [1] for (23), (24), (26)). To prove (27), consider the obvious one-one correspondence between $A(B \oplus C)$ and $AB \oplus AC$. Each element of either can be identified with a couple (a, b) or (a, c) $(a \in A, b \in B, c \in C)$. In both, $(a, b) \geqq (a', b')$ if and only if $a \geqq a'$ and $b \geqq b'$, and similarly for $(a, c) \geqq (a', c')$. In $A(B \oplus C)$, we have $(a, b) \geqq (a', c)$ if and only if $a \geqq a'$; in $AB \oplus AC$, we have $(a, b) \geqq (a', c)$ identically; these give (27).

To show that equality does not hold in (27), note that, in terms of lattice diagrams, we have the situation shown in Fig. 1. As a corollary of (21), (22),

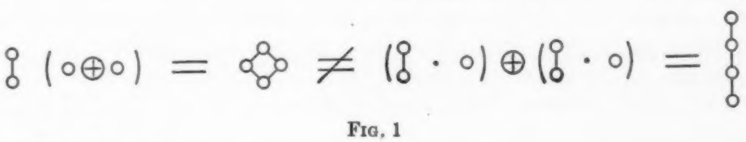

FIG. 1

and (25), we get also

(28) $A = A1 \subset AB$ and $A = A1 \prec AB$, for all B.

This is analogous to (18)–(19) above.

6. **Ordinal or lexicographic multiplication.** The usual definition ([3], p. 78) of the lexicographic product of two ordered sets applies without change to arbitrary "numbers". When so applied, it specializes not only to the usual product of two ordinals, but also (curiously enough) to the product of two cardinals, as usually understood. Nevertheless, we shall not regard it as the correct generalization of ordinary cardinal multiplication, for the reason that it does not, in general, satisfy the identities of cardinal arithmetic.

DEFINITION. By the *ordinal product* of two numbers A and B (in symbols, $A \circ B$) is meant the set of all couples (a, b) $(a \in A, b \in B)$, where $(a, b) \geqq (a', b')$ if and only if $a > a'$ in A *or* $a = a'$ in A and $b \geqq b'$ in B.

In the finite case, we can construct the diagram of $A \circ B$ from the diagrams of A and B as follows. In each circle representing an element a of A, put a replica B_a of B. Then draw segments from all the maximal elements of each B_a and all the minimal elements of each $B_{a'}$ for a' covering[9] a. This rule may be justified by the *covering condition* of

LEMMA 1. *In $A \circ B$, (a, b) covers (a', b') if and only if* (i) *$a = a'$ and b covers b', or* (ii) *b is minimal and b' maximal in B, while a covers a' in A.*

THEOREM 6. *Ordinal multiplication is single-valued, isotone, and associative; it admits* 1 *as an identity, and is semi-distributive on ordinal and cardinal sums alike. Formally,*

(29) $A = B$ *implies* $A \circ C = B \circ C$ *and* $C \circ A = C \circ B$;

(30) $A \subset B$ *implies* $A \circ C \subset B \circ C$ *and* $C \circ A \subset C \circ B$;

(31) $A \prec B$ *implies* $C \circ A \prec C \circ B$;

(32) $A \circ (B \circ C) = (A \circ B) \circ C$ *for all* A, B, C;

(33) $1 \circ A = A \circ 1 = A$ *for all* A;

(34) $(A \oplus B) \circ C = A \circ C \oplus B \circ C$;

(35) $(A + B) \circ C = A \circ C + B \circ C$ *and* $A \circ (B + C) \prec A \circ B + A \circ C$.

Proof. Rules (29)–(31) are immediate if (a, c) is made to correspond to (b, c) if and only if a corresponds to b; the details are easy to check. The proofs of (32)–(33) are also immediate.

In regard to (34)–(35), we use the cardinal one-one correspondence already utilized on proving (26)–(27). First, we have the two equalities. In all cases, the couples (a, c) are ordered as in $A \circ C$ and the couples (b, c) as in $B \circ C$; hence we need merely compare couples (a, c) with couples (b, c'). In (34), it is easy

[9] By the statement *a' covers a*, it is meant that $a' > a$, yet $a' > x > a$ for no x.

to show that always, on both sides, $(a, c) \geqq (b, c')$, while in (35), (a, c) and (b, c') are unrelated in both cases. Then, we have the inequality. In both cases, the couples (a, b) are ordered as in $A \circ B$ and the couples (a, c) as in $A \circ C$; hence we need merely compare (a, c) with (b, c'). But these are never comparable in $A \circ B + A \circ C$; hence the correspondence certainly preserves all inequalities.

Remarks. Ordinal multiplication is thus more closely connected with cardinal than with ordinal addition.

The non-commutativity of ordinal multiplication, known in the infinite case, clearly appears also in the finite case.

One might expect $A \circ (B \oplus C)$ and $A \circ B \oplus A \circ C$ to be related in some way. However, if A, B, and C are finite *and* A is simply ordered, one may show by a little numerical computation that both expressions have the same number of elements *and* of relations; hence, if related by any inequality, they must be isomorphic. Now set A = the ordinal two (or two-element chain), $B = 3$, and $C = 2$. Drawing the appropriate diagrams, we get $A \circ (B \oplus C) \neq A \circ B \oplus A \circ C$.

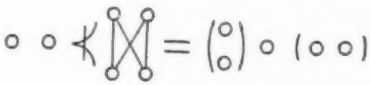

FIG. 2

Surprisingly enough, $A < B$ need *not* imply $A \circ C < B \circ C$; cf. Fig. 2. The same example shows that $A \not< B \circ A$ is possible (for $A = 2$, B the two-element chain as above). On the other hand, it is an immediate corollary of (34)–(35) and the fact that 1 is a subnumber and homonumber of every number, as well as an identity for ordinal multiplication, that

(36) $A \subset A \circ B, B \subset A \circ B, A < A \circ B$, for all A, B.

7. **Cardinal exponentiation.** The usual definition (cf. [3], p. 37) of a cardinal power of a cardinal number is a special case of a general definition, applicable to arbitrary numbers.[10] This general definition is not suggested by any very obvious considerations, but plays an important rôle in lattice theory.

DEFINITION. By the *cardinal power* A^B of the "base" number A raised to the "exponent" B is meant the set of all those functions $f(x)$ with domain B and range in A, which are *isotone* in the sense that $x \geqq y$ in B implies $f(x) \geqq f(y)$ in A. This set is ordered by making $f \geqq g$ in A^B mean that $f(x) \geqq g(x)$ in A for all $x \in B$.

THEOREM 7. *Cardinal exponentiation is single-valued and semi-isotone; it satisfies the usual exponentiation laws with respect to the other cardinal operations and* 1:

[10] This generalization was given in [1]; cf. also [2], p. 13.

(37) $A = B$ *implies* $A^C = B^C$ *and* $C^A = C^B$;

(38) $A \subset B$ *implies* $A^C \subset B^C$;

(39) $A \prec B$ *implies* $C^A \subset C^B$;

$$\left.\begin{aligned} &(40)\ A^{B+C} = A^B A^C, \\ &(41)\ (AB)^C = A^C B^C, \\ &(42)\ (A^B)^C = A^{BC}, \end{aligned}\right\} \textit{for all } A, B, C;$$

(43) $A^1 = A$ *and* $1^A = 1$ *for all* A.

Proof. Law (37) is evident from the abstract nature of the definition of cardinal exponentiation. Since each isotone function from C to A is a fortiori one from C to B if $A \subset B$, and the same inclusion relation holds, (38) is also immediate. Regarding (39), let θ be the given correspondence from B to A. With each g in C^A associate $f = g\phi$, defined by the identity $f(b) = g(b\theta)$. By hypothesis, $b \geqq b'$ implies $b\theta \geqq b'\theta$, hence $g(b\theta) \geqq g(b'\theta)$ and so $f(b) \geqq f(b')$: in short, every $f = g\phi$ is isotone. Finally, if $f^* = g^*\phi$, $f(b) \geqq f^*(b)$ for all b if and only if, for all $a = b\theta$, $g(a) = g(b\theta) \geqq g^*(b\theta) = g^*(a)$, completing the proof. This proves (39). Incidentally, if $A = C =$ the ordinal $1 \oplus 1$ and B is the cardinal number 2, then we have an example where $A \prec B$ yet $A^C = 1 \oplus 1 \oplus 1 \not\prec B^C = 2$. The proofs of (40)–(42) are contained in [1]; those of (43) are trivial. This proves Theorem 7.

It is an immediate corollary of (38) and (43) that

(44) $A \subset A^B$ for all A, B.

In fact, the left-hand side consists of all constant functions occurring on the right-hand side.

8. **Ordinal exponentiation: past methods.** It seems to me that the weakest point of classical transfinite arithmetic comes when ordinal exponentiation is defined.

Consider Cantor's original definition ([3], p. 118)

$$\xi^1 = \xi, \qquad \xi^{\eta+1} = \xi^\eta \xi, \qquad \xi^{\sup \eta} = \sup (\xi^\eta).$$

This inductive definition is *essentially* non-constructive. It is not even equivalent to known constructive definitions, unlike the corresponding inductive definitions of ordinal sums and products

$$\xi + 1 = \xi + 1, \qquad \xi + (\eta + 1) = (\xi + \eta) + 1, \qquad \xi + (\sup \eta) = \sup(\xi + \eta),$$
$$\xi 1 = \xi, \qquad \xi(\eta + 1) = \xi\eta + \xi, \qquad \xi(\sup \eta) = \sup(\xi\eta),$$

which are equivalent to the constructive definitions of §§4, 6.

Second, and this is also not the case with ordinal sums and products, it destroys the otherwise perfect homomorphism from ordinal arithmetic to cardinal arith-

metic. Thus, in the notation of [3], whereas $2^{\omega} = \omega$ is an equation between ordinals, the corresponding cardinal equation $2^{\aleph_0} = \lambda^{\aleph_0}$ is false.

Again, consider Hausdorff's alternative definition: Y^X is the set of all functions from X to Y, ordered according to the first non-zero difference. This is in close conformity with the corresponding cardinal definition, and sufficiently constructive; thus it avoids the defects of Cantor's definition. But it has a peculiar defect all its own: Y^X, although a chain, is not usually an ordinal;[11] thus 2^{ω} is not an ordinal. To be sure, it gives approximately the order-type of the real continuum (actually, that of the Cantor discontinuum), which is a very pretty result. But $2^{2^{\omega}}$ is not even simply ordered, and if we use *it* as an exponent, we get something indescribable. I think it is fair to say that here Hausdorff gave up and defined a "partially ordered set" as one of those pathological things which he got by his construction.

The definition of ordinal exponentiation given below is equivalent to Hausdorff's for both ordinals and cardinals, is constructive, and has the added advantage that, under it, the family of *chains* (simply ordered sets) will at least be closed.

9. **Ordinal exponentiation: new definition.** The change in Hausdorff's definition of ordinal exponentiation which I propose is the following.

DEFINITION. By the *ordinal power* ${}^XY$ is meant the set of all functions f: $y = f(x)$ from X to Y, where $f \geqq g$ means that to each x with $f(x) \ngeq g(x)$ corresponds an $x' > x$ with[12] $f(x') > g(x')$.

This evidently coincides with Hausdorff's definition in the case that X is an ordinal. Its greatest defect is that, although the relation $f \geqq g$ is reflexive and transitive, it is not anti-symmetric: ${}^XY$ is often[13] only a *quasi-ordered* set ([2], p. 7). But it is well known (loc. cit., Theorem 1.2) that such a set becomes a partially ordered set, if $f \equiv g$ is defined to mean $f \geqq g$ and $g \geqq f$. Hence the defect is not essential.

THEOREM 8. *Ordinal exponentiation is single-valued, is slightly isotone, satisfies the law of addition of exponents with respect to both cardinal and ordinal addition and an ordinal semi-associative law of exponentiation. Formally,*

(45) $A = B$ *implies* ${}^CA = {}^CB$ *and* ${}^AC = {}^BC$;

(46) $A \subset B$ *implies* ${}^CA \subset {}^CB$;

[11] In a nutshell, ordinals are not closed under Hausdorff's definition, although cardinals, paradoxically, are. If any definition yielded an ordinal (i.e., failed to have this defect) of the correct power and was constructive (i.e., did not have the defects of Cantor's definition either), it would constructively well-order the continuum. The difficulty of doing this is sufficiently well known (cf. K. Gödel, *The Consistency of the Axiom of Choice and of the Generalized Continuum-hypothesis*, Princeton, 1940).

[12] The idea is that any difference in the values of f and g at x' dominates the values at all points coming afterwards.

[13] Technically, this is the case unless X satisfies the ascending chain condition or Y is totally unordered (a cardinal number).

$$\left.\begin{array}{ll}(47) & {}^{B+C}A = {}^{B}A\,{}^{C}A,\\ (48) & {}^{B\oplus C}A = {}^{B}A \circ {}^{C}A,\\ (49) & {}^{A\circ B}C \prec {}^{A}({}^{B}C),\end{array}\right\}\textit{for all } A, B, C;$$

(50) ${}^{1}A = A$ *and* ${}^{A}1 = A$ *for all* A.

Proof. Law (45) is evident from the abstract nature of the definition of ordinal exponentiation. Again, since each function from C to A is a fortiori one from C to B if $A \subset B$, and inclusion is defined in the same way, (46) holds. Next, we are to prove (47) and (48). The functions f from $B + C$ or $B \oplus C$ to A correspond one-one with the pairs of functions (g, h), one from B to A and the other from C to A. In the cardinal case, $f \geqq f'$ if and only if $g \geqq g'$ and $h \geqq h'$; in the ordinal case, since $a > b$ for all $a \in A$, $b \in B$, $f \geqq f'$ if and only if $g > g'$ or $g = g'$ and $h \geqq h'$. Therefore, the correspondence defines the asserted isomorphism in both cases. Law (50) is trivial. While as for (49), the functions from $A \circ B$ to C assign to each couple (a, b) a value $c = f(a, b)$, hence to each fixed a, a function $f_a(b)$ from B to C; this is simply the usual one-one correspondence from ${}^{A\circ B}C$ to ${}^{A}({}^{B}C)$. But in the first case, $f \ngeqq g$ means that, for some (a, b), (i) $f(a', b') > g(a', b')$ for no $a' > a$, regardless of b', (ii) $f(a, b') > g(a, b')$ for no $b' > b$, and (iii) $f(a, b) \ngeqq g(a, b)$. But now condition (i) implies $f_{a'} > g_{a'}$ for no $a' > a$ while conditions (ii) and (iii) assert that $f_a \ngeqq g_a$; combining, $f \ngeqq g$ in ${}^{A\circ B}C$ implies $f \ngeqq g$ in ${}^{A}({}^{B}C)$. Dualizing, $f \geqq g$ in ${}^{A}({}^{B}C)$ implies $f \geqq g$ in ${}^{A\circ B}C$, which was what we wanted to prove. We note that if A and B satisfy the ascending chain condition, then the equality holds in (49).

10. **Dualization.** In addition to the six binary operations which we have just discussed, there is an important unary operation: that of dualization. This is trivial for cardinals and ordinals, but is very important in most other cases.

Definition. By the *dual* of a number X (in symbols, X^*) is meant the number obtained from X by replacing the inclusion relation in X by its converse (cf. [2], p. 8).

Graphically, this amounts to turning the diagram of X upside down, i.e., to *reversing* the order in X.

Theorem 9. *Dualization is single-valued, involutory, and isotone; it is an isomorphism for all cardinal operations, a dual isomorphism for ordinal addition and multiplication, and a semi-isomorphism on ordinal exponentiation. Formally,*

(51) *if* $X = Y$, *then* $X^* = Y^*$, *while* $(X^*)^* = X$;

(52) *if* $X \subset Y$, *then* $X^* \subset Y^*$; *and if* $X \prec Y$, *then* $X^* \prec Y^*$;

(53) $(X + Y)^* = X^* + Y^*$, $(XY)^* = X^*Y^*$, *and* $(X^Y)^* = X^{*Y^*}$;

(54) $(X \oplus Y)^* = Y^* \oplus X^*$ *and* $(X \circ Y)^* = Y^* \circ X^*$;

(55) $({}^{Y}X)^* = {}^{Y}X^*$.

Proof. All of the above results may be obtained by the reader by appealing to the appropriate definitions. The peculiar non-duality of ordinal exponentiation may help to explain the many vagaries of this operation.

11. **Closure properties.** We shall define below various special classes of numbers, with particular reference to their closure under the various arithmetic relations and operations discussed above. For this, it will be convenient to give names to certain types of closure which appear repeatedly.

First, if S is any set with a binary relation ρ, we shall call a subclass P of S *hereditary* under ρ, when P contains with any element a, all x such that $x\rho a$. (For this terminology, cf. C. Kuratowski, *Topologie*, Warsaw, 1933, p. 29.)

Next, if S is a set with a binary operation $\cdot$, we shall call a subclass P of S *closed* under the operation if it contains $a \cdot b$ whenever it contains a and b. This concept extends also to unary, ternary, and other operations.

Finally, the subclass P will be called a *caste* under the given binary operation, provided it contains $a \cdot b$ if *and only if* it contains both a and b. In the language of genetics, the property of belonging to P is *recessive*.

As examples of recessive properties, we have the following familiar cases: (i) *homogeneity*, for polynomials under multiplication, (ii) case of being a *unit* in a commutative ring (divisor of unity), (iii) case of being *primary* under a given prime ideal, under ideal multiplication.

A "caste" thus defines a congruence relation with two equivalence classes, P and $S - P$; and the complement of a caste has the properties of a (multiplicative) *prime ideal*. What is more important, the property of being a caste under a given operation is a "closure property" in the sense of [2] ("extensionally attainable" in the sense of E. H. Moore). In fact, the caste-closure $\bar{P}$ of a set P is its closure with respect to the operations (i) of including with any element, *all* its ancestors, and (ii) of including with any two elements, their product. For instance, if S is a lattice under the binary operation $\cup$, the caste-closure of any subset is just the *ideal* generated by that subset.

12. **Kinds of numbers.** The following special classes of numbers will be considered below: cardinal numbers, ordinal numbers, chains, lattices, complete lattices, striated numbers, and finite numbers.

A *cardinal* number means a number A such that $x \geqq y$ in A implies $x = y$. With any number X may be associated *its* cardinal number $c(X)$; this is composed of the elements of X, but with a new inclusion relation, which is allowed to subsist only between identical elements.

An *ordinal* number is a number A, every non-void subset S of which has a greatest (first) element s, satisfying $s \geqq x$ for all $x \in S$. This is the usual well-ordering condition. A *chain* is a so-called simply ordered system; a number A such that for any $x, y \in A$ either $x \geqq y$ or $y \geqq x$.

Thus any ordinal number is a chain, while the only number which is both an ordinal and a cardinal is 1, the partially ordered system with a single element.

In applications, a special rôle is played by those numbers which are *lattices* in the sense of [2], that is, numbers A in which, given x and y, there exist a smallest element $x \cup y$ containing both and a largest element $x \cap y$ containing both. Numbers such that this is true not only for two-element subsets, but for arbitrary subsets, are called *complete lattices*.

By a *bounded* number will be meant a number having a least element o and a greatest element i, such that $o \leqq x \leqq i$ for all x. By a *striated* number will be meant a number A which satisfies the Jordan-Dedekind chain condition, in the sense of [2]. More precisely, A will be called striated if and only if each $x \in A$ has a numerical dimension $d[x]$ which is a non-negative integer assuming bounded values, and is such that if x covers y, then $d[x] = d[y] + 1$. By the *length* $d[A]$ of a striated number A is meant the maximum length of a chain in A; this is $d[i] - d[o]$ if A is bounded.

Finally, the notion of a *finite* number will be understood in the obvious way; A is finite if and only if its cardinal number $c(A)$ is finite, that is, if and only if there is no one-one correspondence between A and a proper subset of itself.

13. **Closure properties of classes of numbers.** One can represent in concise tabular form most of the closure properties of the classes of numbers just defined, with respect to the relations and operations of generalized arithmetic. The rows of the table represent the different classes of numbers, and the columns the relations and operations of arithmetic. Thus the entry in the i-th row and j-th column describes the closure properties of the i-th class of numbers with respect to the j-th arithmetic operation or relation.

THEOREM 10. *The following table of closure properties is correct.*

	$\subset$	$\prec$	$+$	$\oplus$	$\cdot$	$\circ$	X^Y	${}^YX$	$*$
Cardinal	H	H	R	O	R	R	BD	BD	C
Ordinal	H	H	O	R	O'	R	O''		
Chain	H	H	O	R	O'	R	O''	R'	C
Lattice			O	C	R		BD	BD	C
Complete lattice			O	C	R	R	BD	$BD?$	C
Bounded		H	O	C	R	R	BD	BD	C
Striated			R	R	R	R	BD'	BD'	C
Finite	H	H	R	R	R	R	R'	R'	C

Explanation. BD (base dominant) means resultant is in subset if and only if base X is. BD' means resultant is in subset if and only if base is, and exponent is finite. C means subset is closed under operation (if operation is unary $*$, this is really the same as R). H means property is hereditary (supra). O means resultant of operation never has property; O' means resultant never has property unless one factor is 1; O'' means power never has property unless base is 1 or base is two-element chain and exponent has property. R means property is recessive (subset in question is caste); R' means property is recessive unless base is 1.

Proof. In almost all cases, the truth of the assertions indicated by the entries is well known or easily verified. We shall only sketch proofs in a few exceptional cases.

For instance, consider the assertion that X^Y is a lattice (complete lattice) if and only if X is. The sufficiency follows since the subset X^Y of isotone functions is a (closed) sublattice of $X^{c(Y)}$; the necessity follows since, if X is not a lattice, the constant functions in X^Y corresponding to meetless or joinless sets of elements of X are meetless respectively joinless in X^Y.

Again, consider the closure property of complete lattices under ordinal multiplication. It is easily shown that

$$\sup(a_\alpha, b_\alpha) = (\sup a_\alpha, \sup_{\alpha\epsilon\sigma} b_\alpha),$$

where σ is the (possibly void) set of all α with $a_\alpha = \sup a_\alpha$. From this it follows easily that complete lattices form a caste under ordinal multiplication.

The closure of chains under ordinal exponentiation is easy to prove. For $f \geqq g$ unless for some x, $f(x) \ngeqq g(x)$, which is the same as $f(x) < g(x)$ if X is a chain, while $f(x') = g(x')$ for all $x' > x$. But in this case, if Y is a chain, it is easy to show that $f < g$.

The author has been unable to prove that the property of being a complete lattice is base dominant under ordinal exponentiation without assuming the ascending chain condition for the exponent Y. In this case, we can define $g = \sup f_\alpha$ as follows. For any $y \epsilon Y$, let A_y denote the subset of all α such that $f_\alpha(y') < g(y')$ for some $y' \epsilon Y$; recursively, we define $g(y) = \sup f_\alpha(y)$ for all $\alpha \notin A_y$.

The other non-obvious cases concern striated lattices. The closure properties under addition are obvious, and

$$d[A + B] = \sup(d[A], d[B]); \qquad d[A \oplus B] = d[A] + d[B]. \tag{56}$$

Under multiplication, we have (a, b) covering (a', b') in AB if and only if $a = a'$ while b covers b' or vice versa. Lemma 1 of §6 takes care of the ordinal case. Together, they give the closure properties of striated numbers, and

$$d[AB] = d[A] + d[B] \quad \text{and} \quad d[A \circ B] = d[A]d[B]. \tag{57}$$

Similarly, under exponentiation, one can verify that

$$d[A^B] = d[A]c(B). \tag{58}$$

The case of ordinal exponentiation is much more complicated, and would take up more space than it is worth.

14. **Special closure properties of lattices.** The closure rules for lattices under ordinal multiplication and exponentiation are so curious as to deserve special mention.

THEOREM 11. *The ordinal product $L \circ M$ of two lattices is a lattice if and only if L is simply ordered or M is bounded.*

Proof. Clearly $(x, y) \cup (x', y')$ must be (x, y), (x', y'), $(x, y \cup y')$, or $(x \cup x', o)$, according as $x > x'$, $x < x'$, $x = x'$, or x and x' are incomparable. These conditions can always be fulfilled if and only if M has an o or the fourth case never arises, i.e., if L is a chain. Dualizing, we get the assertion of the theorem.

Since no l-group is bounded except the trivial l-group having a single element, we get the

COROLLARY.[14] *The ordinal product $G \circ H$ of two l-groups is an l-group if and only if G is simply ordered.*

THEOREM 12. *For ${}^XY$ to be a lattice, it is necessary and sufficient that one of the following three conditions hold:* (i) *Y is a lattice and X a cardinal number,* (ii) *Y is a bounded lattice,* (iii) *Y is a chain and X a semi-root.*

Explanation. A "semi-root" is a partially ordered set in which the elements above each fixed element form a chain.

First, Y must be a lattice, or else not even all pairs of constant functions will have joins and meets. If Y is a lattice, then since

(59) if X is a cardinal number, then ${}^XY = Y^X$,

it is sufficient that X be a cardinal number, by Theorem 10. Again, if Y is a bounded lattice, then ${}^XY$ is a lattice. The only problem is in case $f(x)$ and $g(x)$ are incomparable. In this case, call a *critical value* of x one such that $f(x') = g(x')$ for all $x' > x$, whereas $f(x) \neq g(x)$. Set $h(x) = f(x) \cup g(x)$ at all critical values, and $h(t) = o$ at all points less than critical values. Repeating this process on $(h, f) = h_1$, $(h_1, g) = h_2$, $\cdots$, and using transfinite induction (this is a "sweeping-down process" for critical values), we will arrive ultimately at $f \cup g$, which thus exists. Every h_i is contained in all upper bounds to f and g.

There remains the case that Y is not bounded and X is not a cardinal number. In this case, Y must be a chain, or we could choose $x > x'$ with $f(x)$ and $g(x)$ incomparable but $f(t) = g(t)$ for all $t > x$, whence if $h = f \cup g$, we would have $h(x') = o$, and dually, contradicting the assumption that Y was not bounded. Further, unless X is a semi-root, we can find $x' > x$, $x'' > x$, with x', x'' incomparable. Again, since Y is not bounded, $Y \neq 1$; hence we can choose f and g such that $f(t) = g(t)$ for all $t \neq x', x''$, $f(x') > g(x')$, and $f(x'') < g(x'')$. Again, if $h = f \cup g$, clearly $h(x) = o$; forming $f \cap g$ similarly, Y would have to be bounded, contrary to hypothesis. Hence Y is a chain and X is a semi-root. The sufficiency of these conditions is easy. Given f and g, form $h = f \cup g$ by making $h(x) = f(x)$ unless x is below a critical value t at which $f(t) < g(t)$; at such points, set $h(x) = g(x)$.

[14] This result was implicitly conjectured by Mr. J. C. Abbott while doing graduate work at Harvard University in 1938–1939.

For the notion of l-group, cf. the author's *Lattice-ordered groups*, Annals of Mathematics, vol. 43(1942), pp. 298-331.

Just as before, we have the

COROLLARY. *If Y is an l-group, then ${}^{X}Y$ is an l-group if and only if X is a cardinal number, or X is a semi-root and Y is simply ordered.*

15. **Cardinals and ordinals: special properties.** There are various special properties of ordinal and cardinal numbers which should be mentioned, if only because so many of them hold for more general classes of numbers.

First we note the following more or less trivial properties of the function $c(A)$. We have

(60) $c(A + B) = c(A \oplus B) = c(A) + c(B)$;

(61) $c(AB) = c(A \circ B) = c(A)c(B)$;

(62) $A \prec c(A)$ for all A;

(63) $A^{B} \subset A^{c(B)}$ and ${}^{B}A \prec {}^{c(B)}A$.

Then we have the counterpart of (59),

(64) if A is a cardinal, then $A \circ B = AB$.

The most important special properties of ordinal and cardinal are the following *anti-symmetric* and *comparability* laws.

THEOREM 13. *If A and B are both cardinals or both ordinals, then*

(65) *$A \subset B$ and $B \subset A$ imply $A = B$;*

(66) *either $A \subset B$ or $B \subset A$.*

These laws are well known. We shall see later that the anti-symmetry law is valid also for finite numbers. We may also note without proof that if A and B are both cardinals or both *finite* ordinals, then

(67) $A \prec B$ if and only if $A \subset B$.

An interesting partial extension of this result is the fact that *if B is a chain*, then

(68) $A \prec B$ implies $A \subset B$.

To see this, suppose we take a single representative $b(a)$ from the antecedents of each $a \in A$; the correspondence $b(a) \rightarrow a$ will then be one-one and preserve order; hence it will be an isomorphism.

Another special property of *finite* ordinals is seen in the commutative laws

(69) $A \oplus B = B \oplus A$ and $A \circ B = B \circ A$,

which are valid *for these numbers*. We have further[15]

[15] Cf. S. Sherman, *Some new properties of transfinite ordinals*, Bulletin of the American Mathematical Society, vol. 47(1941), pp. 111–116.

(70) $A \circ (B \oplus C) \subset (A \circ B) \oplus (A \circ C)$, for any ordinals.

Other important special laws for cardinals are the following converses of (18).

THEOREM 14. *If A and B are both cardinals, then*

(71) *$A \subset B$ implies $A = B$ or $A + X = B$ for some X;*

if A and B are both ordinals, then

(72) *$A \subset B$ implies $A = B$ or $A \oplus X = B$ for some X.*

In somewhat the same vein, we may note that if $B \supset A$, and A, B are ordinals, then either $B = Q \circ A$ or $B = Q \circ A \oplus R$ $(R < A)$ for some unique Q, R (right division algorithm). This fact enables one to develop a factorization theory for finite and infinite ordinals (cf. [3]).

16. **Finite numbers: special properties.** When we come to finite numbers, we find first that the anti-symmetric laws hold. More precisely, we have

THEOREM 15. *If A and B are finite numbers, then*

(65) *$A \subset B$ and $B \subset A$ imply $A = B$, and*

(73) *$A \prec B$ and $B \prec A$ imply $A = B$.*

Proof. Law (65) is trivial, since $c(A) = c(B)$. As for (73), first note that since $c(A) \leqq c(B)$ and conversely, A and B have equally many elements. Hence the homomorphisms are one-one. A similar argument shows that they must leave the number of ordered couples (x, y) such that $x \geqq y$ invariant. Hence they are isomorphisms.

More interesting is the study of cancellation laws. These are lost in ordinary transfinite cardinal arithmetic, and half lost in transfinite ordinal arithmetic. The results stated below put them in their true setting.

THEOREM 16. *If A is any finite number, then*

(74) *$A + X = A + Y$ implies $X = Y$.*

If A satisfies the ascending chain condition, then

(75) *$A \oplus X = A \oplus Y$ implies $X = Y$,*

and dually, if A satisfies the descending chain condition,

(75′) *$X \oplus A = Y \oplus A$ implies $X = Y$.*

Proof. As for (74), this follows from the unique decomposition theorem for cardinal addition; an explicit proof can also be given. As for (75)–(75′), by duality, we need only prove (75). If (75) is not true, we can assume (by symmetry) that under the given isomorphism θ from $A \oplus X$ to $A \oplus Y$, $a\theta \in Y$ for some $a \in A$. But this means that for $y = a\theta$, $a\theta^{-1} > y\theta^{-1} = a$, since $a > y$ in $A \oplus Y$ and θ^{-1} is an isomorphism. It follows that A would have to possess

an infinite ascending chain $a < a\theta^{-1} < (a\theta^{-1})^{-1} < \cdots < a\theta^{-n} < \cdots$, contrary to hypothesis.

COROLLARY. *The cancellation law* (75) *is valid for ordinals.*

The author has not attempted to generalize the one-sided cancellation law for ordinal multiplication to all numbers which satisfy the ascending chain condition, but conjectures that this is possible.

If X is finite, or satisfies the ascending chain condition, then ${}^XY$ can be defined to consist of all functions $y = f(x)$ from X into Y, where $f \gneqq g$ means that for some (maximal) $x, f(x) \gneqq g(x)$, while for all $x' > x, f(x) = g(x)$. Using this definition, we can prove the associative law of exponentiation:

(76) ${}^{A \circ B}C = {}^A({}^BC)$, if A and B satisfy the ascending chain condition.

Indeed, the usual one-one correspondence subsists between the elements $c = f(a, b)$ of ${}^{A\circ B}C$ and those $c = f_a(b)$ of ${}^A({}^BC)$. In the former, $f \gneqq g$ means that, for some a, b, (i) $f(a', b') = g(a', b')$ for all $a' > a$ and (i') $f(a, b') = g(a, b')$ for all $b' > b$, while (ii) $f(a, b) \gneqq g(a, b)$. In the latter, $f \gneqq g$ means that, for some a, (i) $f_{a'} = g_{a'}$ for all $a' > a$, while (iii) $f_a \gneqq g_a$. But (iii) means in turn that, for some b, (i') $f_a(b') = g_a(b')$ for all $b' > b$, while (ii) $f_a(b) \gneqq g_a(b)$. The isomorphism can now be read off from the equivalence between the two forms of (i), (i'), (ii).

17. **Bounded numbers.** Bounded numbers also have a number of special properties not true of all numbers. In the first place, we can state the following simple results:

$$\left.\begin{array}{ll}(77) & A \prec A \oplus B, \qquad B \prec A \oplus B \\ (78) & A \prec B \text{ implies } A \circ C \prec B \circ C\end{array}\right\} \text{for bounded numbers.}$$

The reader should have no trouble in proving (77). To prove (78), let θ map B on A as assumed, and let $S(a)$ denote the set of $b \in B$ such that $b\theta = a$. In each $S(a)$, we can (by finite or transfinite induction) choose a maximal non-void set $T(a)$ of incomparable elements—elements such that $x > y$ for no $x, y \in T(a)$. Relative to $T(a)$, each element b of $S(a)$ will fall into just one of the following categories: (i) $b > b'$ for some $b' \in T(a)$, (ii) $b \in T(a)$, (iii) $b < b'$ for some $b' \in T(a)$. Now map (b, c) on $(b\theta, i)$ of $A \circ C$ in case (i), on $(b\theta, c)$ in case (ii), and on $(b\theta, o)$ in case (iii). It may be checked that if $b > b'$, then the image of (b, c) contains that of any (b', c') whether $b\theta > b'\theta$ or $b\theta = b'\theta$; also, that of (b, c) obviously contains that of (b, c') for any $c' \leqq c$.

THEOREM 17. *Any two decompositions of a bounded number into cardinal factors have a common refinement.*[16]

COROLLARY 1. *Any finite bounded number can be factored uniquely into indecomposable ("prime") factors.*

[16] This result is proved in [2], Theorem 2.9.

This implies the following extension of the cancellation law for the multiplication of finite cardinals.

COROLLARY 2. *$AX = AY$ implies $X = Y$ if X, Y, A are finite and bounded.*

This result holds more generally if AX and AY have a finite "center" in the sense of [2]. It suggests the conjectures that if AX and AY are finite, then $AX \subset AY$ may imply $X \subset Y$, and possibly even $AX \prec AY$ may imply $X \prec Y$.

The author knows no example which would prove that the boundedness condition was irredundant in Corollaries 1–2 above. But in (78), if C is the cardinal two, A is one, and B is the ordinal two, we have $A \prec B$ yet $C = A \circ C \nprec B \circ C$; hence the boundedness condition is not redundant.

Regarding cancellation laws, the author conjectures that the following law is valid for finite numbers: $A^C = B^C$ implies $A = B$ if A, B, C are finite. In the case C is a cardinal number, it follows if a unique factorization theorem is known (e.g., if A, B are bounded). In general, it is not even known whether $A^2 = B^2$ implies $A = B$; this has been conjectured by S. Ulam for general abstract systems. It is certain that $C^A = C^B$ does *not* imply $A = B$; thus if C is a cardinal number, then $C^A = C$ for all lattices, and indeed for all numbers A not representable as cardinal sums.

18. **Special interpretations with lattices.** In the case of lattices, it is natural to replace the relation $A \subset B$ by the stronger relation $A \subset {}^*B$, meaning that A is a *sublattice* of B. Similarly, it is natural to replace the relation $A \prec B$ by the stronger relation $A \prec {}^*B$, meaning that A is a *lattice-homomorphic* image of B. We have the following cross-connection:

(79) $A \prec {}^*B$ implies $A \subset B$ if B is a finite lattice.

For, the correspondence $a \to \sup_{b\theta = a} b$ is an isomorphism. If B is distributive, we even have $A \subset {}^*B$.

Many of the isotonicity and duality laws (cf. (10), (11), (18), (19), (21), (22), (52), etc.) proved above for the relations $\subset$ and $\prec$ are valid also for the stronger relations $\subset^*$ and $\prec^*$. In particular, we have

(80) $A \subset B$ implies $C^A \prec {}^*C^B$.

At least in case C is the ordinal two, we have (see Garrett Birkhoff, *Rings of sets*, this Journal, vol. 3(1937), p. 454) the curious counterpart that $A \prec B$ implies $C^A \subset {}^*C^B$—and in fact, that all sublattices of C^B may be obtained in this way.

We also have (and in the finite case this follows from laws (79) and (80))

(81) $A \subset B$ implies that $C^A \subset C^B$ if C is complete.

Proof. Given f from A to C, define $g = f\theta$ by

$$g(b) = \sup_{a \leq b} f(a) \qquad \text{for any } b \in B.$$

Clearly g is isotone; and, if f is isotone, makes $g(b) = f(b)$ for all $b \in A$. Hence g is an *extension* of f, and the correspondence $f \to g$ is one-one; it is evidently isotone.

The author has no counterexample to (81) for numbers not complete lattices, but the possibility of extension does not exist in all cases.

19. **Applications.** We now come to the main argument in favor of a broader attitude towards the six arithmetic operations and dualization: the fact that in the wider context of general partially ordered sets, many new applications of these operations are found. Let us study this situation, first as to the cardinal operations, and then as to the ordinal operations.

For the sake of comparison, it should first be stressed that the cardinal operations really have very few applications in traditional transfinite arithmetic.[17] The operations of addition and multiplication are actually trivial, since the sum of any two infinite numbers, like their product, is simply the *larger* of the two summands (multiplicands). The remaining operation, that of exponentiation, is primarily useful in constructing from $\aleph_0$ the only known infinite cardinals, including $c = 2^{\aleph_0}$.

In contrast to this slim array of classical applications, there are known at least nine distinct applications of our extended cardinal arithmetic to questions of lattice theory,[18] especially to the theory of Boolean algebras and distributive lattices. Besides, they may be applied to topology ([2], p. 15). If M and N are any two abstract complexes, then $M + N$ represents their topological sum and MN their topological (or "Cartesian") product. While if M represents any subdivision of a manifold without boundary, then M^* represents the dual of the subdivision.

Again, it is fair to ask, what applications of the *ordinal* operations of traditional transfinite arithmetic are known? Transfinite induction should not be included, as it does not involve addition, multiplication, or exponentiation.[19] The operations of addition and multiplication are primarily useful in that they afford a neat notation for countable ordinals; the same is true of Cantor's inductively defined operation of exponentiation. These ordinals are used in such places as the Baire classification of functions. Also, with Hausdorff's exponentiation operation, 2^ω gives the Cantor discontinuum.

Equally important, it seems to me, are the uses of these operations to partially ordered systems which are not well-ordered. In the first place, ordinal multi-

[17] The distinction between $\aleph_0$ and c, although fundamental in modern analysis, is independent of the *operations* of cardinal arithmetic.

[18] These are listed in [1], §3, and will not be repeated here. They mainly concern exponentiation, although the operation of cardinal multiplication (forming the direct product) is also of fundamental importance in lattice theory.

[19] It would be a mistake to minimize the importance of transfinite induction. However, it should be noted that transfinite sequences are being replaced by directed sets in topology (cf. J. W. Tukey's *Convergence and Uniformity in Topology*, Princeton, 1940), and that transfinite induction is being replaced by the Lemma of Zorn in many other connections.

plication is useful in the construction and description of non-Archimedean ordered groups,[20] which are so basic in modern valuation theory. Again, the most general vector lattice with finite basis can be described as ${}^{X}R$, where R is the real number system and X is the most general "semi-root".[21] Further, the lattice of l-ideals of ${}^{X}R$ is B^{X}, where B denotes the ordinal two, thus establishing a curious connection between cardinal and ordinal exponentiation. Finally, the most general vector lattice can be built up from R by repeated cardinal and ordinal multiplication, to form such vector lattices as $(R \circ (RR))\,(R \circ R)$.

BIBLIOGRAPHY

1. G. BIRKHOFF, *An extended arithmetic*, this Journal, vol. 3(1937), pp. 311–316.
2. G. BIRKHOFF, *Lattice Theory*, New York, 1940.
3. F. HAUSDORFF, *Grundzüge der Mengenlehre*, first edition, Leipzig, 1914.
4. M. H. STONE, *The Theory of Real Functions*, Ann Arbor, Michigan, 1940.

HARVARD UNIVERSITY.

[20] Cf., for example, H. Hahn, *Ueber die nichtarchimedische Grössensysteme*, S.-B. Wiener Akad., Math.-Nat. Klasse, Abt. IIa, vol. 116(1907), pp. 601–653.

[21] For the facts stated here, cf. the author's paper *Lattice-ordered groups* and the doctoral thesis of Mr. Murray Mannos.

MAXIMAL FIELDS WITH VALUATIONS

By Irving Kaplansky

1. **Introduction.** A field with a valuation is said to be *maximal* if it possesses no proper immediate extensions, i.e., if every extension of the field must enlarge either the value group or the residue class field. This definition is due to F. K. Schmidt, but was first published by Krull ([4], p. 191). In the same paper Krull succeeded in proving that any field with a valuation possesses at least one immediate maximal extension and that any field of formal power series is maximal in its natural valuation. These facts led Krull to propound the following two queries.

(1) Is the immediate maximal extension of a field uniquely determined?

(2) If a maximal field K has the same characteristic as its residue class field, is K necessarily a power series field?

These two closely related questions form the central problem of this investigation. The answer to the first is obtained in §3 (Theorem 5), as follows. The immediate maximal extension is always unique if the residue class field has characteristic ∞; but if the latter has characteristic p, then a pair of conditions which we have labelled "hypothesis A" must be satisfied. It is then not difficult to obtain the answer to the second question in §4. In fact, with the same hypothesis, the answer is again affirmative, provided factor sets are admitted in the construction of the power series field (Theorem 6). Granted an additional hypothesis, it is furthermore possible to dispense with factor sets (Theorem 8). In §5, examples are given to show that the conclusions of the preceding theorems may fail if hypothesis A is not fulfilled.

The notion of pseudo-convergence, borrowed from Ostrowski ([9], p. 368), appears to be a natural tool for investigations of maximality, and it is employed consistently throughout the paper. The reason for this is to be found in Theorem 4, which shows that pseudo-convergence provides us with an *intrinsic* characterization of maximality.

2. **Pseudo-convergence and maximality.** Throughout this section K will always denote a field with a valuation V on an ordered Abelian group Γ, B its valuation ring, and $\mathfrak{K}$ its residue class field.[1]

Definition. A well-ordered set $\{a_\rho\}$ of elements of K, without a last element, is said to be *pseudo-convergent* if

$$V(a_\sigma - a_\rho) < V(a_\tau - a_\sigma) \tag{1}$$

Received November 3, 1941; presented to the American Mathematical Society, February 22, 1941. The author wishes to express his thanks to Professor MacLane for his assistance in the preparation of this paper.

[1] For these definitions, cf. [4] and [7].

for all $\rho < \sigma < \tau$.[2]

LEMMA 1.[3] *If $\{a_\rho\}$ is pseudo-convergent, then either*
(i) *$V(a_\rho) < V(a_\sigma)$ for all $\rho < \sigma$, or*
(ii) *$V(a_\rho) = V(a_\sigma)$ from some point on, i.e., for all $\rho, \sigma \geqq$ some ordinal λ.*

Proof. Suppose that (i) does not hold, i.e., that $V(a_\rho) \geqq V(a_\sigma)$ for some $\rho < \sigma$. Then $V(a_\tau)$ must equal $V(a_\sigma)$ for all $\tau > \sigma$. For, if not, we would have

$$V(a_\tau - a_\sigma) = \min [V(a_\sigma), V(a_\tau)] \leqq V(a_\sigma),$$

while $V(a_\sigma - a_\rho) \geqq V(a_\sigma)$, so that the inequality (1) could not possibly hold.

LEMMA 2. *If $\{a_\rho\}$ is pseudo-convergent, then $V(a_\sigma - a_\rho) = V(a_{\rho+1} - a_\rho)$ for all $\rho < \sigma$.*

Proof. We may assume $\sigma > \rho + 1$. From the inequality

$$V(a_{\rho+1} - a_\rho) < V(a_\sigma - a_{\rho+1}),$$

and the identity

$$a_\sigma - a_\rho = (a_\sigma - a_{\rho+1}) + (a_{\rho+1} - a_\rho),$$

we deduce that

$$\begin{aligned} V(a_\sigma - a_\rho) &= \min [V(a_\sigma - a_{\rho+1}), V(a_{\rho+1} - a_\rho)] \\ &= V(a_{\rho+1} - a_\rho). \end{aligned}$$

As a consequence of Lemma 2, we can unambiguously introduce the abbreviation γ_ρ for $V(a_\sigma - a_\rho)$ $(\rho < \sigma)$. We note that $\{\gamma_\rho\}$ is a monotone increasing set of elements of Γ.

DEFINITION.[4] An element x of K is said to be a *limit* of the pseudo-convergent set $\{a_\rho\}$ if $V(x - a_\rho) = \gamma_\rho$ for all ρ.

DEFINITION. The set of all elements y of K such that $Vy > \gamma_\rho$ for all ρ forms an (integral or fractional) ideal in the valuation ring B; this ideal we call the *breadth*[5] of $\{a_\rho\}$.

The limit of a pseudo-convergent set is by no means unique; however, given one limit, it is easy to describe the totality of limits.

LEMMA 3. *Let $\{a_\rho\}$ be pseudo-convergent with breadth $\mathfrak{A}$, and let x be a limit of $\{a_\rho\}$. Then an element is a limit of $\{a_\rho\}$ if and only if it is of the form $x + y$, with $y \in \mathfrak{A}$.*

Proof. If z is any other limit, it follows from

$$x - z = (x - a_\rho) - (z - a_\rho)$$

[2] Cf. [9], p. 368. The inequalities here read in the opposite sense because Ostrowski uses an exponential valuation.

[3] Cf. [9], p. 369.

[4] This definition does not always coincide with Ostrowski's, [9], p. 375.

[5] A translation of "Breite", [9], p. 368.

that $V(x - z) > \gamma_\rho$ for all ρ, whence $x - z$ lies in $\mathfrak{A}$. Conversely, if $y \in \mathfrak{A}$, then

$$Vy > \gamma_\rho = V(x - a_\rho),$$

and so $V(x + y - a_\rho) = \gamma_\rho$, whence $x + y$ is a limit of $\{a_\rho\}$.

Let the field L be an extension of K, with a valuation which is an extension of V. If the value group and residue class field of L coincide with Γ and $\mathfrak{K}$, respectively, we say that L is an *immediate* extension of K. If K admits no proper immediate extensions, K is said to be *maximal*. It will now be our object to prove that maximality is equivalent to the possession of a limit by every pseudo-convergent set; half of this equivalence is obtained in the following theorem.

THEOREM 1. *Let L be an immediate extension of K. Then any element in L but not in K is a limit of a pseudo-convergent set of elements of K, without a limit in K.*

Proof.[6] Let z be an element in L but not in K, and let S denote the totality of values $V(z - a)$, with a in K. Certainly S does not include the symbol ∞. Further, S cannot have a greatest member γ. For, suppose $V(z - g) = \gamma$, $g \in K$; let $c \in K$ have value γ, and let $d \in K$ be a representative of the residue class of $(z - g)/c$. Then $V(z - g - cd) > \gamma$, where $g + cd \in K$, a contradiction.

From the set S select a well-ordered cofinal subset[7] $\{\alpha_\rho\}$; since S has no greatest member, $\{\alpha_\rho\}$ cannot have a last term. Choose elements $a_\rho \in K$ with $V(z - a_\rho) = \alpha_\rho$. The identity

$$a_\sigma - a_\rho = (z - a_\rho) - (z - a_\sigma),$$

together with the inequality

$$V(z - a_\rho) < V(z - a_\sigma) \qquad (\rho < \sigma),$$

then imply

$$V(a_\sigma - a_\rho) = V(z - a_\rho) \qquad (\rho < \sigma), \tag{2}$$

whence $\{a_\rho\}$ is pseudo-convergent with z as limit.

Suppose that $\{a_\rho\}$ had the further limit z_1 in K. Then, by Lemma 3,

$$V(z - z_1) > V(a_\sigma - a_\rho) \qquad (\rho < \sigma).$$

Combining this with (2) and using the fact that $\{a_\rho\}$ has no last member, we obtain

$$V(z - z_1) > V(z - a_\rho) = \alpha_\rho$$

for all ρ; and this is a contradiction, since $\{\alpha_\rho\}$ is cofinal in S.

Next, we must show that if some pseudo-convergent set $\{a_\rho\}$ in K lacks a limit, then K is not maximal. This will be accomplished by adjoining to K a

[6] It is perhaps worth remarking that Theorem 1 and the preceding lemmas do not depend on the commutativity of either $\mathfrak{K}$ or Γ.

[7] [3], p. 129.

limit of $\{a_\rho\}$ and then proving that the resulting extension is immediate. Since we shall later be interested in questions of uniqueness, Theorems 2 and 3 will also include some preliminary results on uniqueness.

We borrow from Ostrowski the following two lemmas ([9], p. 371, IV and III). His proofs are readily adapted for the more general case under consideration here.

LEMMA 4. *Let $\beta_1, \cdots, \beta_m$ be any elements of an ordered Abelian group Γ, and further let $\{\gamma_\rho\}$ be a well-ordered, monotone increasing set of elements of Γ, without a last element. Let $t_1, \cdots, t_m$ be distinct positive integers. Then there will exist an ordinal μ and an integer k $(1 \leqq k \leqq m)$ such that*

$$\beta_i + t_i\gamma_\rho > \beta_k + t_k\gamma_\rho$$

for all $i \neq k$ and $\rho > \mu$.

LEMMA 5. *If $\{a_\rho\}$ is pseudo-convergent in K, and $f(x)$ is a polynomial with coefficients in K, then $\{f(a_\rho)\}$ is ultimately pseudo-convergent.*[8]

By combining Lemmas 1 and 5 we can make a useful deduction concerning the set $\{Vf(a_\rho)\}$, namely that for all sufficiently large ρ and σ either

$$Vf(a_\rho) = Vf(a_\sigma) \tag{3}$$

or

$$Vf(a_\rho) < Vf(a_\sigma) \qquad (\rho < \sigma) \tag{4}$$

must hold. The distinction between these two cases will persist throughout the discussion, and for convenience we introduce the following definitions.

DEFINITIONS. A pseudo-convergent set $\{a_\rho\}$ in K is said to be of *transcendental type* (with respect to K) if (3) holds for every polynomial $f(x)$ with coefficients in K; if, on the other hand, (4) holds for at least one polynomial $f(x)$, we shall say that $\{a_\rho\}$ is of *algebraic type*.

THEOREM 2. *If there is a pseudo-convergent set $\{a_\rho\}$ of transcendental type in K, without a limit in K, then there exists an immediate transcendental extension $K(z)$ of K. The valuation of $K(z)$ can be specifically defined as follows: for any polynomial $f(z)$ with coefficients in K we define $Vf(z)$ to be the fixed value which $Vf(a_\rho)$ ultimately assumes. In the resulting valuation, $K(z)$ is an immediate extension of K, and z is a limit of $\{a_\rho\}$.*

Conversely, if $K(u)$ is any extension of K, with a valuation which is an extension of V such that u is a limit of $\{a_\rho\}$, then $K(u)$ and $K(z)$ are analytically equivalent over K.[9]

[8] It is to be noted that Ostrowski's pseudo-convergence need only hold from some point on.

[9] By an analytic equivalence over K we mean a value preserving isomorphism which is the identity on K.

Proof.[10] We must first verify that the above definition actually defines a valuation of $K(z)$, i.e., we must show that

$$V[g(z)h(z)] = Vg(z) + Vh(z) \tag{5}$$

and

$$V[g(z) + h(z)] \geqq \min [Vg(z), Vh(z)] \tag{6}$$

for all rational functions $g(z)$ and $h(z)$. But the truth of (5) and (6) follows at once from the truth of the corresponding equations with z replaced by a_ρ.

Next, we wish to show that, with this valuation, $K(z)$ is an immediate extension of K. By definition, $Vf(z) = Vf(a_\rho)$ for large ρ, so there is clearly no extension of the value group. To prove the same for the residue class field, it will suffice to take any polynomial $f(z)$ with $Vf(z) = 0$, and find an element $b \in K$ with $V[f(z) - b] > 0$, for then $f(z)$ and b will lie in the same residue class. Since $\{f(a_\rho)\}$ is ultimately pseudo-convergent, we have

$$V[f(a_\tau) - f(a_\sigma)] > V[f(a_\sigma) - f(a_\rho)] \geqq 0$$

for sufficiently large $\rho < \sigma < \tau$. But, by definition,

$$V[f(z) - f(a_\sigma)] = V[f(a_\tau) - f(a_\sigma)]$$

for large τ. Therefore, $V[f(z) - f(a_\sigma)] > 0$ so that $f(z)$ and $f(a_\sigma)$ lie in the same residue class.

To show that z is a limit of $\{a_\rho\}$, we observe that

$$V(z - a_\rho) = V(a_\sigma - a_\rho) \qquad (\rho < \sigma) \tag{7}$$

for large ρ. An application of Lemma 1 to the pseudo-convergent set $\{z - a_\rho\}$ yields that $\{V(z - a_\rho)\}$ is monotone increasing. Then from the identity

$$(z - a_\rho) - (z - a_\sigma) = a_\sigma - a_\rho,$$

we obtain (7) for all ρ.

It remains to prove the final statement of Theorem 2. Regardless of the characteristic of K, it is possible to form a Taylor expansion for a polynomial $f(u)$ of degree m:

$$f(u) - f(a_\rho) = (u - a_\rho)f_1(a_\rho) + \cdots + (u - a_\rho)^m f_m(a_\rho), \tag{8}$$

where $f_k(u)$ may be thought of as replacing the formal expression $f^{(k)}(u)/k!$. (See, for example, [1], p. 165, Ex. 2, or [2].) By hypothesis it is possible to cut into $\{a_\rho\}$ so far that the values of $f\{a_\rho\}, f_1(a_\rho), \cdots, f_m(a_\rho)$ are all independent of ρ. We shall suppose that this has been done, and let us write β_i for $Vf_i(a_\rho)$ $(i = 1, \cdots, m)$. We apply Lemma 4 with $t_i = i$ $(i = 1, \cdots, m)$ and $\gamma_\rho = V(u - a_\rho)$. Since

$$\beta_i + i\gamma_\rho = V[(u - a_\rho)^i f_i(a_\rho)],$$

[10] Similar results are proved in [4], p. 194 and [9], p. 374.

it follows that for sufficiently large ρ some one of the terms

$$(u - a_\rho)^i f_i(a_\rho) \qquad (i = 1, \cdots, m)$$

has smaller value than all the others. This means that the value of the right member of (8) increases monotonically with ρ for large ρ. Since $Vf(a_\rho)$ is fixed, this is possible only if $Vf(u) = Vf(a_\rho)$ for large ρ, which in turn implies that $Vf(u) = Vf(z)$. We have obtained an explicit analytical equivalence over K between the fields $K(u)$ and $K(z)$.

THEOREM 3. *If $\{a_\rho\}$ is a pseudo-convergent set of algebraic type in K, without a limit in K, then there exists an immediate algebraic extension $K(z)$ of K, which can be explicitly obtained as follows. Among the polynomials $f(x)$ for which* (4) *holds, choose one of least degree n—say $q(x)$. Let z be a root of $q(x) = 0$, and for any polynomial $f(z)$ of degree less than n, define $Vf(z)$ to be the fixed value which $Vf(a_\rho)$ ultimately assumes. In the resulting valuation, $K(z)$ is an immediate extension of K, and z is a limit of $\{a_\rho\}$.*

Conversely, if u is a root of $q(x) = 0$, and if $K(u)$ has a valuation which is an extension of V such that u is a limit of $\{a_\rho\}$, then $K(u)$ and $K(z)$ are analytically equivalent over K.

Proof. First, it is necessary to remark that the polynomial $q(x)$ is irreducible and of degree $\geqq 2$. For, if $q(x) = b(x - c)$, then $V(c - a_\rho)$ increases monotonically for large ρ. But $\{c - a_\rho\}$ is pseudo-convergent; by Lemma 1, $V(c - a_\rho)$ increases monotonically for all ρ, whence it follows that

$$V(c - a_\rho) = V(a_\sigma - a_\rho) \qquad (\rho < \sigma),$$

and c is a limit of $\{a_\rho\}$, contrary to hypothesis. Again, if $q(x) = q_1(x)q_2(x)$, where q_1 and q_2 are polynomials of degree less than n, then $V[q_1(a_\rho)q_2(a_\rho)]$ increases monotonically for large ρ; the same must, therefore, hold for either $Vq_1(a_\rho)$ or $Vq_2(a_\rho)$, contradicting the minimal choice of $q(x)$.

The remainder of the proof, with one exception, is a duplication of the proof of Theorem 2, the discussion being, of course, confined to polynomials of degree less than n. The one exceptional point is the proof of the multiplicative character of the valuation of $K(z)$, and this proof we shall now give.

$K(z)$ consists of polynomials in z of degree less than n, with coefficients in K. The product $h(z)$ of two such polynomials $f(z)$ and $g(z)$ is defined by an equation of the form

$$f(z)g(z) = h(z) + k(z)q(z).$$

We have, for all ρ,

$$f(a_\rho)g(a_\rho) - h(a_\rho) = k(a_\rho)q(a_\rho). \tag{9}$$

Now, for large ρ, the value of the right member of (9) increases monotonically, while $V[f(a_\rho)g(a_\rho)]$ and $Vh(a_\rho)$ are fixed for large ρ. This is possible only if

$$Vh(a_\rho) = V[f(a_\rho)g(a_\rho)]$$

for large ρ, whence, by the definition of V on $K(z)$,

$$Vh(z) = Vf(z) + Vg(z),$$

as desired.

Upon combining Theorems 1, 2, and 3, we obtain

THEOREM 4. *A field with a valuation is maximal if and only if it contains a limit for each of its pseudo-convergent sets.*

3. **Uniqueness of the maximal extension.** It was proved by Krull ([4], Th. 24, p. 191) that any field with a valuation possesses at least one immediate maximal extension. It is natural to inquire whether this extension is uniquely determined. More precisely, if N and N' are two immediate maximal extensions of K, we ask whether there exists between N and N' an analytical equivalence over K.

It is first of all clear from Theorems 2 and 3 that N or N' can be obtained from K by a transfinite series of adjunctions of limits of pseudo-convergent sets. If we can demonstrate that each of these adjunctions takes place in a unique fashion, we shall have obtained an affirmative answer to the question of uniqueness. In the case of transcendental pseudo-convergent sets, uniqueness is already assured us by Theorem 2. It only remains to examine sets of algebraic type, and here, as will appear, uniqueness can indeed fail.

By the use of Theorem 3, we can reformulate the question as follows. Suppose the pseudo-convergent set $\{a_\rho\}$ is of algebraic type in K, and let $q(x)$ be a polynomial of least degree such that $Vq(a_\rho)$ ultimately increases monotonically. Let N be any immediate maximal extension of K.

(*) Does N contain a limit of $\{a_\rho\}$ which is also a root of $q(x) = 0$?

It is now clear that the answer to the question of the uniqueness of the maximal extension hinges entirely on the answer to the question (*).

We shall adopt the following fixed notation for the discussion. Let the degree of q be n. Let q_i denote the i-th formal derivative of q. Cut into $\{a_\rho\}$ sufficiently far so that $Vq_i(a_\rho)$ $(i = 1, \cdots, n)$ is independent of ρ, equal, say, to β_i.[11] Denote $V(a_\sigma - a_\rho)$ $(\rho < \sigma)$ by γ_ρ. Finally, let $\mathfrak{K}$, the residue class field of K, have characteristic p. We treat explicitly only the case where p is finite, but as a matter of fact the proof can be read equally well for the case $p = \infty$; it is only necessary to replace throughout all powers of p by unity.

First, we prove a simple number-theoretic lemma.

LEMMA 6. *If p is prime, and r is a positive integer prime to p, $r > 1$, then $\binom{p^t r}{p^t}$ is prime to p, for any integer $t \geqq 0$.*

Proof.

$$\binom{p^t r}{p^t} = \frac{p^t r(p^t r - 1) \cdots (p^t r - p^t + 1)}{p^t(p^t - 1) \cdots 1}.$$

[11] The fact that some of the β's may be infinite does not vitiate any of the arguments.

In the numerator of this fraction, the first factor is divisible by precisely p^t, while the remaining ones are not divisible by p^t. Hence, for every factor m occurring in the numerator, the factor $m - p^t(r-1)$ which occurs in the denominator will be divisible by p to precisely the same power. This gives the desired result.

LEMMA 7. *If $i = p^t$, $j = p^t r$ with $r > 1$, $(r, p) = 1$, then*

$$\beta_i + i\gamma_\rho < \beta_j + j\gamma_\rho$$

for all sufficiently large ρ.

Proof. We form a Taylor expansion for $q_i(a_\rho)$. In doing so it is necessary to introduce certain binomial coefficients ([2], p. 226).

$$\begin{aligned} q_i(a_\sigma) - q_i(a_\rho) &= (i+1)(a_\sigma - a_\rho)q_{i+1}(a_\rho) + \cdots \\ &+ \binom{j}{i}(a_\sigma - a_\rho)^{j-i}q_j(a_\rho) + \cdots + \binom{n}{i}(a_\sigma - a_\rho)^{n-i}q_n(a_\rho). \end{aligned} \tag{10}$$

Consider the right member of (10) with $\rho < \sigma$. By Lemma 4, there will be among these terms precisely one of least value, provided ρ is sufficiently large. The value of this term must then equal the value of the left member of (10), which in turn is not less than β_i. It follows that the term

$$\binom{j}{i}(a_\sigma - a_\rho)^{j-i}q_j(a_\rho),$$

occurring in (10), must also have value not less than β_i. But, by Lemma 6, $\binom{j}{i}$ has value zero. Therefore,

$$\beta_i \leqq (j-i)\gamma_\rho + \beta_j$$

and the result follows at once from the fact that $\{\gamma_\rho\}$ is monotone increasing.

LEMMA 8. *There is an integer h, which is a power of p, such that for all sufficiently large ρ*

$$\beta_i + i\gamma_\rho > \beta_h + h\gamma_\rho \qquad (i \neq h) \tag{11}$$

and

$$Vq(a_\rho) = \beta_h + h\gamma_\rho. \tag{12}$$

Proof. Consider

$$q(a_\sigma) = q(a_\rho) + (a_\sigma - a_\rho)q_1(a_\rho) + \cdots + (a_\sigma - a_\rho)^n q_n(a_\rho) \tag{13}$$

with $\rho < \sigma$. Applying Lemma 4 to the terms on the right of (13) other than $q(a_\rho)$, we find that for large ρ there is precisely one of them, say $(a_\sigma - a_\rho)^h q_h(a_\rho)$, of least value; this proves (11). Now, keeping ρ fixed and varying σ, we observe that $Vq(a_\sigma)$ increases monotonically. This is possible only if

$$\begin{aligned} Vq(a_\rho) &= V[(a_\sigma - a_\rho)^h q_h(a_\rho)] \\ &= \beta_h + h\gamma_\rho. \end{aligned}$$

That h is a power of p is an immediate consequence of Lemma 7.

Throughout the succeeding discussion, we shall reserve the letter h for the integer occurring in Lemma 8.

LEMMA 9. *If y is a limit of $\{a_\rho\}$, then*

$$Vq(y) > \beta_h + h\gamma_\rho \tag{14}$$

for all ρ, and

$$Vq_i(y) = \beta_i \qquad (i = 1, \cdots, n). \tag{15}$$

Proof. We have

$$q(y) = q(a_\rho) + (y - a_\rho)\, q_1(a_\rho) + \cdots + (y - a_\rho)^n q_n(a_\rho), \tag{16}$$

where, by definition, $V(y - a_\rho) = \gamma_\rho$. By Lemma 8, the terms of least value on the right of (16) are $q(a_\rho)$ and $(y - a_\rho)^h q_h(a_\rho)$. Therefore, $Vq(y) \geqq Vq(a_\rho)$, and, in view of the monotone increasing character of $\{\gamma_\rho\}$, we obtain (14). To prove (15) we form the Taylor expansion

$$q_i(y) - q_i(a_\rho) = (i + 1)(y - a_\rho)q_{i+1}(a_\rho) + \cdots + \binom{n}{i}(y - a_\rho)^{n-i}q_n(a_\rho). \tag{17}$$

By Lemma 4, the value of the right member of (17) increases for large ρ; hence, $Vq_i(y) = Vq_i(a_\rho)$.

The following result is not required in the present connection, but, for convenience, the proof will be given at this point.

LEMMA 10. *Suppose the value group Γ is Archimedean, and suppose further that the breadth of $\{a_\rho\}$ is not the zero ideal. Let*

$$q^*(x) = \sum_{i=p^u} (x - a_\theta)^i q_i(a_\theta),$$

the summation ranging as indicated only over powers of p. Then for sufficiently large θ, and $\rho > \theta$,

$$Vq^*(a_\rho) = Vq(a_\rho) = \beta_h + h\gamma_\rho\,. \tag{18}$$

Proof. Suppose i is not a power of p, and let k be the highest power of p dividing i. By Lemma 7, for large ρ,

$$\beta_i + i\gamma_\rho > \beta_k + k\gamma_\rho\,. \tag{19}$$

The hypothesis that the breadth of $\{a_\rho\}$ is not the zero ideal means that the real numbers γ_ρ approach a finite limit S. From (19), a fortiori,

$$\beta_i + iS > \beta_k + kS.$$

It follows that we can choose a fixed μ so large that

$$\beta_i + i\gamma_\mu > \beta_k + k\gamma_\rho \tag{20}$$

for all ρ. Thus, for every i not a power of p, there is a corresponding μ such that (20) holds. Let θ be any ordinal exceeding all these μ's. Then

$$\beta_i + i\gamma_\theta > \beta_k + k\gamma_\rho \geqq \beta_h + h\gamma_\rho \tag{21}$$

holds for every i not a power of p. Next, we write

$$q(a_\rho) = \sum_{i=0}^{n} (a_\rho - a_\theta)^i q_i(a_\theta). \tag{22}$$

Since, by Lemma 7, $Vq(a_\rho) = \beta_h + h\gamma_\rho$, it follows from (21) that we can eject from (22) the terms for which i is not a power of p without altering the value of the right hand member; this yields (18) at once.

We are now able to obtain our principal result on the uniqueness of the immediate maximal extension. In the event that the residue class field $\mathfrak{K}$ has finite characteristic p, the requisite hypothesis is contained in the following two statements:

(1) Any equation of the form

$$x^{p^n} + a_1 x^{p^{n-1}} + \cdots + a_{n-1} x^p + a_n x + a_{n+1} = 0,$$

with coefficients in $\mathfrak{K}$, has a root in $\mathfrak{K}$.[12]

(2) The value group Γ satisfies $\Gamma = p\Gamma$.

For convenience, we shall refer to this pair of conditions as "hypothesis A". If the characteristic of $\mathfrak{K}$ is infinite, we shall further agree that hypothesis A is vacuous. The proof of the following theorem can then be read for this case in the light of our previous remark that all p-th powers are to be replaced by unity.[13]

THEOREM 5. *Let the field K have a valuation with value group Γ and residue class field $\mathfrak{K}$, such that $\mathfrak{K}$ and Γ satisfy hypothesis* A. *Then the immediate maximal extension of K is uniquely determined up to analytical equivalence over K.*

Proof. As we remarked above, it follows from Theorems 2 and 3 that we need only prove the following statement: if $\{a_\rho\}$ is pseudo-convergent of algebraic type in K with $q(x)$ for a minimal polynomial, and if N is any immediate maximal extension of K, then N contains a limit of $\{a_\rho\}$ which is also a root of $q(x) = 0$. This is done by a transfinite approximation, for which purpose it is convenient first to prove the following lemma. (The symbols γ_ρ, β_i and h are used as defined above.)

LEMMA 11. *If for some limit $t \in N$ of $\{a_\rho\}$ we have $Vq(t) = \alpha$, we can obtain the better approximation $Vq(t^*) > \alpha$, where $t^* \in N$ is a limit of $\{a_\rho\}$ such that*

$$V(t^* - t) = \max_{i=p^u} (\alpha - \beta_i)/i,$$

i ranging as indicated over the powers of p $(1 \leqq i \leqq n)$.

[12] Concerning this rather unusual hypothesis we wish to remark that it definitely falls short of algebraic closure. An example is provided by taking the Galois field of p elements $(p > 2)$ and closing it off with respect to extensions of odd degree.

[13] In fact, the whole discussion could be greatly shortened if we were interested in this case only. In particular, the transfinite approximation in Theorem 5 could be replaced by a simple application of the Hensel-Rychlik theorem.

Proof. Write

$$\delta = \max (\alpha - \beta_i)/i, \tag{23}$$

the range of i being the powers of p, as it will be throughout the proof. Taking $i = h$ in (23) and using Lemma 9, we obtain

$$\delta > (\beta_h + h\gamma_\rho - \beta_h)/h = \gamma_\rho \tag{24}$$

for all ρ. Let $k \in N$ be any element of value δ. (This is possible, as hypothesis A implies $\delta \in \Gamma$.) For any $z \in N$,

$$q(t + kz)/q(t) = \sum_{j=0}^{n} k^j z^j q_j(t)/q(t). \tag{25}$$

In the polynomial (25) the coefficient of z^j has value $j\delta + \beta_j - \alpha$. If j is a power of p, we have

$$j\delta + \beta_j - \alpha \geqq 0 \tag{26}$$

by (23). If j is not a power of p, and i is the highest power of p dividing j, then

$$j\delta + \beta_j - \alpha > i\delta + \beta_i - \alpha \geqq 0$$

by Lemma 7, (24) and (26). Taking these facts together, we observe that if we replace each coefficient in (25) by its residue class, we obtain a polynomial with coefficients in $\mathfrak{K}$, say $\bar{F}(z)$, of precisely the type used in hypothesis A. Hence, $\mathfrak{K}$ contains a root $\bar{z}_1$ of $\bar{F}(z) = 0$. If $z_1 \in N$ is any representative of the residue class $\bar{z}_1$, we then have $V[q(t + kz_1)/q(t)] > 0$ or $Vq(t + kz_1) > \alpha$. Also, by the choice of k, $V(kz_1) = \delta$. By (24), kz_1 lies in the breadth of $\{a_\rho\}$, whence by Lemma 3, $t + kz_1$ is a limit of $\{a_\rho\}$. With the choice of $t^* = t + kz_1$, we have therefore proved Lemma 11.

We now resume the proof of Theorem 5. We are going to select a transfinite set of elements $\{t_\mu\}$ of N such that:

(1) each t_μ is a limit of $\{a_\rho\}$;

(2) if $Vq(t_\mu) = \alpha_\mu$, then $\alpha_\mu < \alpha_\nu$ $(\mu < \nu)$;

(3) $V(t_\nu - t_\mu) = \max (\alpha_\mu - \beta_i)/i$ $(\mu < \nu)$, the range of i again being the powers of p.

Let us first observe that the proof of Theorem 5 can then be easily completed; for, a cardinal number consideration shows that the choice of the t's must terminate with the appearance of an element $t_\zeta \in N$, which is a limit of $\{a_\rho\}$, and for which $Vq(t_\zeta) = \infty$, or $q(t_\zeta) = 0$. For t_1, we choose any limit of $\{a_\rho\}$ in N (there is at least one by Theorem 4), and suppose t_μ has been chosen for all $\mu < \lambda$ so as to satisfy (1), (2), and (3) for $\mu < \nu < \lambda$.

(i) λ a limit number. Then (2) and (3) imply

$$V(t_\nu - t_\mu) < V(t_\theta - t_\nu) \qquad (\mu < \nu < \theta < \lambda),$$

showing that $\{t_\mu\}_{\mu<\lambda}$ is pseudo-convergent. Let $t_\lambda \in N$ be any limit of $\{t_\mu\}_{\mu<\lambda}$.

As an immediate consequence of the definition of limit, we have

$$V(t_\lambda - t_\mu) = \max (\alpha_\mu - \beta_i)/i \qquad (\mu < \lambda). \tag{27}$$

From (27) and Lemma 9, as in (24), we have $V(t_\lambda - t_\mu) > \gamma_\rho$, so that $t_\lambda - t_\mu$ is in the breadth of $\{a_\rho\}$, and t_λ is a limit of $\{a_\rho\}$ by Lemma 3. Finally, we must prove

$$Vq(t_\lambda) > \alpha_\mu \qquad (\mu < \lambda). \tag{28}$$

We write

$$q(t_\lambda) = q(t_\mu) + (t_\lambda - t_\mu)q_1(t_\mu) + \cdots + (t_\lambda - t_\mu)^n q_n(t_\mu). \tag{29}$$

For j a power of p, by Lemma 9,

$$V[(t_\lambda - t_\mu)^j q_j(t_\mu)] = \beta_j + j \max (\alpha_\mu - \beta_i)/i \geqq \alpha_\mu ; \tag{30}$$

while for j not a power of p, (30) follows a fortiori from Lemma 7. Applying these facts to (29), we obtain $Vq(t_\lambda) \geqq \alpha_\mu$, which suffices to prove (28).

(ii) λ not a limit number. Here $t_{\lambda-1}$ is given, and by Lemma 11, we can find a limit t_λ of $\{a_\rho\}$ such that

$$Vq(t_\lambda) > \alpha_{\lambda-1} \tag{31}$$

and

$$V(t_\lambda - t_{\lambda-1}) = \max (\alpha_{\lambda-1} - \beta_i)/i. \tag{32}$$

From (31) and (32), respectively, (28) and (27) readily follow. With this the induction is complete.

4. The structure of maximal fields.

The results obtained in §3 will now enable us to obtain explicit theorems on the structure of maximal fields and their representations as power series fields.

If $\mathfrak{K}$ is any field and Γ is any ordered Abelian group, the set of all formal series

$$\sum a_\rho t^{\alpha_\rho} \qquad (a_\rho \in \mathfrak{K}, \alpha_\rho \in \Gamma, \{\alpha_\rho\} \text{ well-ordered})$$

form a field, when addition and multiplication are defined in the usual formal fashion. This field we may denote by $\mathfrak{K}(t^\Gamma)$.[14] In $\mathfrak{K}(t^\Gamma)$ we can introduce a valuation V by setting

$$V(\sum a_\rho t^{\alpha_\rho}) = \alpha_1 \qquad (a_1 \neq 0).$$

Krull has proved that in this valuation $\mathfrak{K}(t^\Gamma)$ is maximal ([4], p. 193).

We now pose the converse query: is a maximal field K, with value group Γ and residue class field $\mathfrak{K}$, analytically isomorphic to $\mathfrak{K}(t^\Gamma)$? As the first step in obtaining such a representation, we must find a subfield M of K which can serve as the coefficient field. Let H denote the homomorphism mapping every

[14] For power series fields cf. [4], [8], [10].

$a \epsilon K$ with $Va \geqq 0$ into its residue class, and mapping every $a \epsilon K$ with $Va < 0$ into ∞.[15] Then the property we desire for M is represented by the equation $H(M) = \mathfrak{K}$.

LEMMA 12. *Let K have the same characteristic as its residue class field $\mathfrak{K}$, and suppose K is algebraically perfect, and satisfies the Hensel-Rychlik theorem.*[16] *Then K possesses a subfield M with $H(M) = \mathfrak{K}$.*

Proof. The proof requires only a slight amplification of Lemma 2, [7]. If P and $\mathfrak{P}$ denote the prime subfields of K and $\mathfrak{K}$, respectively, we necessarily have $H(P) = \mathfrak{P}$. We build up M by successive adjunctions as in MacLane's proof, and the only point that needs further investigation is the case of an inseparable algebraic extension. Suppose then that we have $H(N) = \mathfrak{N}$, and we wish to obtain an extension of N corresponding to $N(\bar{a}^{1/p})$, $\bar{a} \epsilon \mathfrak{N}$. Let $a \epsilon N$ be the representative of $\bar{a}$. By hypothesis $a^{1/p} \epsilon K$, and $N(a^{1/p})$ provides the desired extension.

Next we must obtain a set of elements playing the rôle of the elements $\{t^\alpha\}$ in a power series field. It will, however, in general be necessary to admit a factor set for the multiplication of these elements.

LEMMA 13. *With the same hypothesis as in Lemma 12, and with a fixed choice of the field M, there exists a set $\{t^\alpha\}$ in K, with $Vt^\alpha = \alpha$ for every $\alpha \epsilon \Gamma$, and with*

$$t^\alpha t^\beta = c_{\alpha,\beta} t^{\alpha+\beta} \qquad (c_{\alpha,\beta} \epsilon M),$$

where $c_{\alpha,\beta}$ is a factor set.

Proof. By well-known methods, it is possible to choose a rationally independent basis $\{\zeta_\rho\}$ for Γ, i.e., a set $\{\zeta_\rho\}$ such that every $\alpha \epsilon \Gamma$ has a unique representation as a sum of ζ's with rational coefficients. For t^ζ we choose any element of value ζ; and if α is a sum of ζ's with integral coefficients, we choose for t^α the product of the corresponding elements t^ζ raised to the appropriate powers. Suppose finally that α is of the form

$$\alpha = r_1\zeta_1 + \cdots + r_m\zeta_m$$

with not all the r_i integral. Let r be the L.C.M. of the denominators of $r_1, \cdots, r_m$; then $t^{r\alpha}$ has already been assigned. We shall now show that it is possible to select an element t^α such that

$$(t^\alpha)^r = at^{r\alpha} \qquad (a \epsilon M). \tag{33}$$

Since K is perfect, it suffices to take the case $(r, p) = 1$. Let $z \epsilon K$ have value α, and let a be the M-representative of $z^r/t^{r\alpha}$. Then $at^{r\alpha}/z^r$ lies in the same residue class as 1, and, by the Hensel-Rychlik theorem, has an r-th root in K.

[15] A detailed statement of the connection between H and V is given in [7].

[16] In [4], p. 178, this theorem is proved on the hypothesis of completeness, a weaker condition than maximality.

Therefore, $at^{r\alpha}$ also has an r-th root, and this is our choice for t^{α}. For any other $\beta \in \Gamma$ we will have the similar equations:

$$(t^{\beta})^{s} = bt^{s\beta} \qquad (b \in M), \tag{34}$$

$$(t^{\alpha+\beta})^{u} \ ct^{u(\alpha+\beta)} \qquad (c \in M). \tag{35}$$

From (33), (34), and (35) we obtain

$$(t^{\alpha}t^{\beta}/t^{\alpha+\beta})^{rsu} = a^{su}b^{ru}/c^{rs} \in M.$$

Since M is a coefficient field, it follows from this that $t^{\alpha}t^{\beta}/t^{\alpha+\beta} \in M$, as desired.

Before we can obtain our structure theorem, it will be necessary to prove the following result.

LEMMA 14. *Let N be a maximal field of characteristic p, with value group Γ, and residue class field $\mathfrak{K}$, and suppose $\mathfrak{K}$ is perfect and $\Gamma = p\Gamma$. Then N is perfect.*

Proof. Suppose that, on the contrary, $a \in N$ has no p-th root in N. We construct the extension $M = N(a^{1/p})$. Because N is complete in Krull's sense, the valuation of N extends to M in the following unique manner.[17] Any b in M but not in N will satisfy an irreducible equation of the form

$$x^{p} + c_1x^{p-1} + \cdots + c_p = 0 \qquad (c_i \in N)$$

and to b we assign the value $V(c_p)/p$. Plainly this involves no extension of Γ. Furthermore, since any element in M is a p-th root of some element in N, it follows readily that there is no extension of $\mathfrak{K}$. Hence M is an immediate extension of N, contrary to the hypothesis that N is maximal.

The corresponding lemma for the characteristic unequal case will also be needed, but here a stronger hypothesis must be made.

LEMMA 15. *Let N be maximal and of infinite characteristic, while the residue class field $\mathfrak{K}$ has characteristic p, and suppose Γ and $\mathfrak{K}$ satisfy hypothesis* A (*cf. Theorem* 5). *Then every element of N has a p-th root in N.*

Proof. We employ a transfinite approximation. Since this is carried out in virtually the same fashion as in Theorem 5, we shall here merely summarize the method.

First, it suffices to prove $a^{1/p} \in N$ for $Va = 0$. We make an inductive choice of elements $\{t_\rho\}$, with $V(t_\rho^p - a) = \alpha_\rho$, such that

$$\alpha_\rho < \alpha_\sigma \qquad (\rho < \sigma),$$

$$V(t_\sigma - t_\rho) = \max(\alpha_\rho/p, \alpha_\rho - Vp) \qquad (\rho < \sigma).$$

For t_1 we choose any element of N with $V(t_1^p - a) > 0$. (There is such an element, since hypothesis A implies that the residue class of a has a p-th root.) Suppose we have chosen t_ρ for all $\rho < \lambda$. If λ is not a limit number, it follows just as in Lemma 11 that there exists t_λ with $V(t_\lambda^p - a) > \alpha_{\lambda-1}$, and $V(t_\lambda - t_{\lambda-1}) = \max(\alpha_{\lambda-1}/p, \alpha_{\lambda-1} - Vp)$. If λ is a limit number, then $\{t_\rho\}_{\rho<\lambda}$ is pseudo-con-

[17] [4], p. 180.

vergent; for t_λ we then choose any limit of $\{t_\rho\}_{\rho<\lambda}$. When the approximation terminates, we obtain a p-th root of a in N, as desired.

We are now able to prove our first structure theorem. For brevity let us denote by $\mathfrak{K}(t^\Gamma, c_{\alpha,\beta})$ the power series field in which multiplication takes place according to the rule: $t^\alpha t^\beta = c_{\alpha,\beta}t^{\alpha+\beta} (c_{\alpha,\beta} \in \mathfrak{K})$.

THEOREM 6. *Let the maximal field N, with value group Γ and residue class field $\mathfrak{K}$, have the same characteristic as K; and suppose $\mathfrak{K}$ and Γ satisfy hypothesis* A. *Then N is analytically isomorphic to a power series field $\mathfrak{K}(t^\Gamma, c_{\alpha,\beta})$.*

Proof. By Lemma 14, N is algebraically perfect. We are then able to apply Lemmas 12 and 13, obtaining a coefficient field M which we may identify with $\mathfrak{K}$ and a set of representatives $\{u^\alpha\}$ with

$$u^\alpha u^\beta = c_{\alpha,\beta}u^{\alpha+\beta} \qquad (c_{\alpha,\beta} \in \mathfrak{K}).$$

Form the subfield K of N obtained by adjoining to $\mathfrak{K}$ all the elements u^α, and let K' denote the analogous subfield of $\mathfrak{K}(t^\Gamma, c_{\alpha,\beta})$, i.e., the field obtained by adjoining to $\mathfrak{K}$ all elements t^α. Let T be the natural map of K' on K, i.e., T is the identity on $\mathfrak{K}$ and $Tt^\alpha = u^\alpha$. Plainly T is a homomorphism; but, moreover, T preserves values, and so must actually be an analytic isomorphism. Now N and $\mathfrak{K}(t^\Gamma, c_{\alpha,\beta})$ are immediate maximal extensions of K and K', respectively. By Theorem 5 and our hypothesis, N and $\mathfrak{K}(t^\Gamma, c_{\alpha,\beta})$ are analytically isomorphic.

It is natural to inquire in what circumstances the factor set occurring in Theorem 6 can be dispensed with. For this purpose we require an extension theorem of wider scope than Theorem 5. The investigation also yields a uniqueness theorem for the characteristic unequal case, as will appear in Theorem 8.

THEOREM 7.[18] *Let the field K have value group Γ and residue class field $\mathfrak{K}$, and let the two maximal extensions L and L' of K have value group Δ and residue class field $\mathfrak{L}$, which may be proper extensions of Γ and $\mathfrak{K}$. Then if Δ and $\mathfrak{L}$ satisfy hypothesis* A, *and if every element of $\mathfrak{L}$ has an n-th root in $\mathfrak{L}$ for all n, L and L' are analytically equivalent over K.*

Proof. It will suffice to prove that L and L' contain analytically equivalent subfields N and N' with value group Δ and residue class field $\mathfrak{L}$; for then L and L' are analytically equivalent by Theorem 5. We will build up N and N' through a transfinite succession of fields paralleling adjunctions in $\mathfrak{L}/\mathfrak{K}$ and Δ/Γ, and it suffices to consider the case of a single adjunction.

Residue class field adjunction. A transcendental or separable algebraic extension is handled exactly as in [6], Theorem 3; to treat an inseparable algebraic extension, we observe that, by hypothesis and Lemmas 14 and 15, every element in L has a p-th root in L; so we are again able to use MacLane's method.

Value group adjunction. Consider an extension $\Gamma(\alpha)$ of Γ. If α is rationally

[18] This is the non-discrete analogue of [6], Theorem 3, and [5].

independent of Γ, simply let $a \in L$, $a' \in L'$ be any elements of value α. Then for any polynomial $f(x) = \sum c_i x^i$ with coefficients in K we have

$$Vf(a) = Vf(a') = \min V(c_i a^i), \tag{36}$$

showing that $K(a)$ and $K(a')$ are analytically equivalent. If α is rationally dependent on Γ, we have $n\alpha = \beta$ for some $\beta \in \Gamma$. Let $b \in K$ have value β; we wish to show that b has an n-th root in L. Using Lemmas 14 and 15 we reduce the consideration to the case $(n, p) = 1$. Let $z \in L$ have value α. By hypothesis, the residue class of b/z^n has an n-th root in L. By the Hensel-Rychlik theorem, b/z^n has an n-th root in L, whence b has an n-th root, say a, in L. Likewise b has an n-th root a' in L'. Then (36) holds for polynomials of degree less than n, showing again that $K(a)$ and $K(a')$ are analytically equivalent.

We now obtain our second structure theorem.

THEOREM 8. *Let the maximal field K have value group Γ and residue class field $\mathfrak{R}$, and suppose that Γ and $\mathfrak{R}$ satisfy hypothesis* A, *and that every element of $\mathfrak{R}$ has an n-th root in $\mathfrak{R}$, for all n. Then K is uniquely determined, up to analytic isomorphism by $\mathfrak{R}$, Γ, its characteristic, and in the characteristic unequal case, Vp.*

Proof. Let P be the prime subfield of K. The valuation of P is uniquely determined by the given data, up to analytic isomorphism. Theorem 8 then follows from Theorem 7.

COROLLARY. *In the equal characteristic case, every field with a valuation is analytically isomorphic to a subfield of a suitable power series field.*

5. **Counter-examples.** We will show by an example that without hypothesis A the conclusion of Theorem 5 may fail.[19]

Let $\mathfrak{R}$ be a field of characteristic p, and let Γ be the additive group of all rational numbers. Let K be the subfield of the power series field $\mathfrak{R}(t^\Gamma)$ obtained by adjoining to $\mathfrak{R}$ all the elements t^α; K is then the field of all quotients of linear expressions in the t's, with coefficients in $\mathfrak{R}$. Consider the pseudo-convergent sequence $\{a_i\}$:

$$a_i = t^{-1/2} + t^{-1/4} + \cdots + t^{-1/2^i}.$$

We wish first to show that $\{a_i\}$ is of transcendental type in K. Now the breadth of $\{a_i\}$ is in fact precisely the valuation ring of K; at any rate, it is not the zero ideal. If $\{a_i\}$ were of algebraic type in K, then by Lemma 10 there would exist elements $c_j \in K$ $(j = 0, \cdots, n + 1)$ such that the value of

$$z_i = c_0 a_i^{p^n} + c_1 a_i^{p^{n-1}} + \cdots + c_{n-1} a_i^p + c_n a_i + c_{n+1}$$

increases monotonically for large i. We can suppose without loss of generality that $Vc_j \geqq 0$ $(0 \leqq j \leqq n)$, $Vc_j = 0$ for at least one j in the same range, and

[19] The principle upon which this example is constructed is the same as in the first of the counter-examples of [7].

further that all the c's are actually polynomials in the t's. We now imagine z_i multiplied out in full, and consider the portion of z_i consisting of terms t^α with $\alpha < 0$; call this portion w_i. Suppose c_j begins with the term $d_j \in \mathfrak{R}$ (d_j may be zero); then the contribution of $c_j a_i^{p^{n-j}}$ to w_i will consist of $d_j a_i^{p^{n-j}}$ together with some other terms, the latter being fixed for large i. Moreover, for different j's, the contributions $d_j a_i^{p^{n-j}}$ are elementwise distinct, at any rate if $p \neq 2$. It must then be the case that, for large i, w_i will once and for all contain a fixed term of least value, and this statement is incompatible with the previous assertion that Vz_i increases monotonically for large i.

Now suppose that hypothesis A is violated because the equation

$$g(x) = x^{p^n} + b_1 x^{p^{n-1}} + \cdots + b_n x = b, \tag{37}$$

with coefficients in $\mathfrak{R}$, has no root in $\mathfrak{R}$. The formal series

$$a = t^{-1/2} + t^{-1/4} + t^{-1/8} + \cdots$$

is a limit of $\{a_i\}$, and likewise $g(a)$ is a limit of $\{g(a_i)\}$, which along with $\{a_i\}$ is of transcendental type in K. By Lemma 3, $g(a) + b$ is also a limit of $\{g(a_i)\}$. Hence, by Theorem 2, $g(a)$ and $g(a) + b$ are both transcendental over K and the mapping $g(a) \to g(a) + b$ provides an analytic automorphism of the field $K[g(a)] = L$, say. The adjunction of a to L, which is plainly immediate, can be paralleled by the corresponding adjunction of a root a' of the equation

$$g(x) = g(a) + b. \tag{38}$$

Let N and N' be any immediate maximal extensions of $L(a)$ and $L(a')$, respectively. Then there cannot exist any analytic isomorphism between N and N' which leaves L elementwise fixed. For, if there were such an isomorphism, then N', like N, would contain a root of $g(x) = g(a)$. But N' already contains a root of (38). Hence N' would contain a root of (37); any such root would necessarily have value zero, and, taking residue classes, we would obtain a root of (37) in $\mathfrak{R}$, contrary to hypothesis.

We have thus shown that the violation of hypothesis A may entail the existence of inequivalent maximal extensions. But this example still leaves another question unanswered, for it might nevertheless be true that all the immediate maximal extensions of the above field K are analytically isomorphic to $\mathfrak{R}(t^\Gamma)$. (This actually occurs in the discrete finite rank case; MacLane [7] gives examples of non-unique extensions, but Schilling [10] has proved that all such fields are power series fields.) However, in the non-discrete case which we are considering, uniqueness fails even in this broader sense. To show this, a somewhat more complicated example is needed.

We shall use the same notation as in the preceding example and in addition the abbreviations $p^n = q$ and $t^{1/2^n p^n} = w_n$. Suppose that hypothesis A is violated, this time by the fact that the element $b \in \mathfrak{R}$ has no p-th root in $\mathfrak{R}$. We adjoin to the field K the following elements in turn: a, $(a + bt)^{1/p} = u_1$,

$(u_1 + bw_1)^{1/p} = u_2, \cdots, (u_n + bw_n)^{1/p} = u_{n+1}$, etc. To assign a valuation to these extensions, we argue as before. First, we find that

$$u_n^q = a + bt + b^p t^{1/2} + \cdots + b^{q/p} t^{1/2^{n-1}}. \tag{39}$$

Since a is a limit of $\{a_i\}$, so is u_n^q by Lemma 3. Hence, u_n is a limit of $\{a_i^{1/q}\}$ and, again by Lemma 3, so is $u_n + bw_n$. Also $\{a_i^{1/q}\}$, along with $\{a_i\}$, is of transcendental type in K. By Theorem 2, the mapping $u_n + bw_n \rightarrow a^{1/q}$ provides an analytical equivalence over K between the fields $K(u_n)$ and $K(a^{1/q})$. Then the extension $K(u_{n+1})$ of $K(u_n)$ can be given a valuation paralleling that of $K(a^{1/pq})$, and in this valuation $K(u_{n+1})$ is an immediate extension of $K(u_n)$.

Let N be any immediate maximal extension of the field $K(a, u_1, u_2, \cdots)$. We shall prove that N is not analytically isomorphic to a power series field.

First we need the following elementary observation. If in a power series field M of characteristic p we have a pseudo-convergent set $\{a_\rho\}$ such that each a_ρ has a p^m-th root in M, then M contains some limit of $\{a_\rho\}$ which also has a p^m-th root in M. To prove this it suffices to construct the power series y which agrees with a_ρ for all terms of value less than $V(a_{\rho+1} - a_\rho)$. Then y is a limit of $\{a_\rho\}$, and, since p^m-th root extraction goes termwise, y has a p^m-th root in M.

Now, in our case, a_i has a p^m-th root in N for all m. If N were a power series field, it would, therefore, contain a limit z of $\{a_i\}$, with a p^m-th root in N for all m. By Lemma 3, $V(z - a) \geqq 0$. Write $z = a + c + z_1$, where $c \in \mathfrak{K}$, $Vz_1 > 0$, say, for definiteness, $Vz_1 > 1/2^{n-1}$. Now $a + c + z_1$ is a p^n-th (or q-th) power in N; together with (39) this implies that $c - b^{q/p} t^{1/2^{n-1}} +$ terms of higher value is a q-th power in N. This means that the residue class of c has a q-th root, whence, since $\mathfrak{K}$ is a coefficient field, c has a q-th root in $\mathfrak{K}$. Subtracting c, and repeating the argument, we obtain that $b^{q/p}$ has a q-th root in $\mathfrak{K}$, i.e., b has a p-th root in $\mathfrak{K}$, contrary to our initial assumption.

Remark. This example is easily duplicated if, instead of $\mathfrak{K}$, it is Γ that is imperfect, i.e., if $\Gamma \neq p\Gamma$. But the author has not succeeded in constructing a counter-example on the assumption that the general equation (37) lacks a root. Thus the possibility remains open that a weaker condition than hypothesis A will suffice to ensure that a maximal field in the equal characteristic case is a power series.

Bibliography

1. A. A. Albert, *Modern Higher Algebra*, Chicago, 1937.
2. H. Hasse and F. K. Schmidt, *Noch eine Begründung der Theorie der höheren Differentialquotienten*, Journal für Mathematik, vol. 177(1937), pp. 215–237.
3. F. Hausdorff, *Mengenlehre*, first edition, Berlin, 1914.
4. W. Krull, *Allgemeine Bewertungstheorie*, Journal für Mathematik, vol. 167(1932), pp. 160–196.
5. S. MacLane, *Note on the relative structure of p-adic fields*, Annals of Mathematics, vol. 41(1940), pp. 751–753.

6. S. MacLane, *Subfields and automorphism groups of p-adic fields*, Annals of Mathematics, vol. 40(1939), pp. 423–442.
7. S. MacLane, *The uniqueness of the power series representation of certain fields with valuations*, Annals of Mathematics, vol. 39(1938), pp. 370–382.
8. S. MacLane, *The universality of formal power series fields*, Bulletin of the American Mathematical Society, vol. 45(1939), pp. 888–890.
9. A. Ostrowski, *Untersuchungen zur arithmetischen Theorie der Körper*, Mathematische Zeitschrift, vol. 39(1935), pp. 269–404.
10. O. F. G. Schilling, *Arithmetic in fields of formal power series in several variables*, Annals of Mathematics, vol. 38(1937), pp. 551–576.

Harvard University.

ALGEBRAIC PROPERTIES OF CERTAIN MATRICES OVER A RING

By Neal H. McCoy

1. **Introduction.** A considerable part of recent work in the theory of matrices has been devoted to the study of matrices with elements in some domain more general than the field of complex numbers, which is so prominent in early papers on this subject. By restricting the domain to be a suitably chosen field or ring, different parts of the classical theory have been generalized, or analogues found, in various ways. In the case in which the elements are from a non-commutative ring, two quite different approaches to the subject have been used. In one of these, there is no attempt to introduce the concept of *determinant*, but sufficiently strong divisibility conditions are assumed in order to carry over certain parts of the theory. In the other approach, which is used in this paper, the class of matrices considered is restricted in such a way that determinants, having many of the familiar properties of ordinary determinants, can be defined for the matrices under consideration. Reference here can be made to the work of E. H. Moore ([6],[1] Chap. II) on Hermitian matrices in what he calls a "number system of type B". As the present investigation was inspired by this work of Moore or, more precisely, by the similarity between his theorems and known theorems about arbitrary matrices over a commutative ring, we pause to describe briefly the class of matrices to which Moore's theory is directly applicable.

For the moment, let $\mathfrak{T}$ be a ring with unit element 1, in which the equation $2x = 1$ has a unique solution and in which there is defined an anti-automorphism or *involution* $a \rightarrow \bar{a}$. Thus

$$\overline{a + b} = \bar{a} + \bar{b}, \qquad \overline{ab} = \bar{b}\bar{a}, \qquad \bar{\bar{a}} = a.$$

We require also that the elements a of $\mathfrak{T}$ such that $a = \bar{a}$, the so-called *symmetric elements*, shall be in the center of $\mathfrak{T}$. Such a ring may be called an *involutorial ring*. If $A = (a_{ij})$ is a square matrix with elements in $\mathfrak{T}$, and $a_{ij} = \bar{a}_{ji}$, then A is said to be a *Hermitian matrix*. Now a "number system of type B", as defined by Moore, is a special instance of an involutorial ring and is in fact either a commutative field, of characteristic other than 2, with $a = \bar{a}$, or a quadratic field over the field of symmetric elements or a generalized quaternion algebra over this field. However, Jacobson has pointed out in [2] that Moore's definition of determinant and many of his results remain valid for Hermitian matrices over any involutorial ring.

Although many of Moore's theorems coincide in statement with known theorems about matrices over a commutative ring, the published proofs are quite different. We shall begin by showing how to unify these two cases, at least to a

Received November 8, 1941.

[1] Numbers in square brackets refer to the bibliography at the end of the paper.

certain extent. If $\mathfrak{R}$ is an arbitrary ring with unit element, we consider not the class $\mathfrak{R}_n$ of all matrices of order n over $\mathfrak{R}$ but a certain subclass $\mathfrak{R}_n'$ of $\mathfrak{R}_n$ consisting of those matrices of $\mathfrak{R}_n$ whose elements satisfy certain weakened commutativity relations to be defined precisely in §2. It is for elements of $\mathfrak{R}_n'$ that we shall be able to carry over much of Moore's theory. It will be found that if $\mathfrak{R}$ is a commutative ring, then $\mathfrak{R}_n' = \mathfrak{R}_n$, while if $\mathfrak{R}$ is an involutorial ring, $\mathfrak{R}_n'$ contains all Hermitian matrices of $\mathfrak{R}_n$. Thus, both these cases appear as special instances of the general theory.

In §2, we present the notation to be used and discuss the class $\mathfrak{R}_n'$ in some detail. In the next section we introduce, following Moore, the concept of adj A. If A is a Hermitian matrix over an involutorial ring, it follows at once from the definition that adj A is also Hermitian. It is only in extending this result to elements of $\mathfrak{R}_n'$ that we need to make any essential extension of Moore's work. However, we shall show in Theorem 2 that if $A \in \mathfrak{R}_n'$, then adj $A \in \mathfrak{R}_n'$. The proof is rather long and detailed, but it is necessary to establish the theorem in order to make use of the concept of determinant of adj A. Further properties of adj A and related results are presented in §4. Up to this point, our work consists roughly of showing that much of Moore's theory is valid for elements of $\mathfrak{R}_n'$. The rest of the paper is not so directly a generalization of Moore's work.

Let A be a fixed element of $\mathfrak{R}_n$, and $\mathfrak{R}[\lambda]$ the ring of polynomials in the indeterminate λ, with coefficients in $\mathfrak{R}$. The set of all elements

$$g(\lambda) = a_0 + a_1\lambda + \cdots + a_m\lambda^m$$

of $\mathfrak{R}[\lambda]$ such that

$$g_r(A) = a_0 + a_1A + \cdots + a_mA^m = 0$$

is a left ideal $\mathfrak{m}_l$ in $\mathfrak{R}[\lambda]$. In §5, we shall show how to characterize this ideal under the assumption that A is an element of $\mathfrak{R}_n'$. Naturally, the ideal $\mathfrak{m}_r$ defined in an analogous way is also considered. If $\mathfrak{C}$ is the center of $\mathfrak{R}$, the ideal $\mathfrak{m}$ in $\mathfrak{C}[\lambda]$ of those elements $h(\lambda)$ of $\mathfrak{C}[\lambda]$ such that $h(A) = 0$ is of primary importance. This ideal $\mathfrak{m}$ we call the *minimum ideal* of A to emphasize the fact that, in case $\mathfrak{R}$ is a field, it is the principal ideal generated by the minimum function of A (cf. [5]). An important property of the minimum ideal of an element of $\mathfrak{R}_n'$ is also to be found in §5. Then, in §6, we show that a recent theorem of Ostrowski [7] can be generalized to the case under discussion. This theorem furnishes, for a set of matrices, a partial analogue of the notion of minimum ideal of a single matrix.[2]

We are unable to establish for arbitrary elements of $\mathfrak{R}_n'$ many theorems known to be true in the commutative case. The remainder of the paper is devoted to the proof of several of these theorems under further strong restrictions. We introduce in §7, in a natural way, the notion of a *quaternion ring*

[2] For the case in which R is a commutative ring, this result has already been obtained in [4].

and thereafter limit ourselves either to the case of Hermitian matrices over a quaternion ring or to arbitrary matrices over a commutative ring. The theorems obtained in §8 are largely concerned with polynomials in a matrix, and are theorems which have already been established in [3] and [5] for the commutative case.

2. **Notation and general remarks.** The following notation will be used throughout. By $\mathfrak{R}$ we denote an arbitrary ring with unit element 1 and center $\mathfrak{C}$. The complete matric ring of order n over $\mathfrak{R}$ or $\mathfrak{C}$ will be denoted by $\mathfrak{R}_n$ or $\mathfrak{C}_n$ respectively. We shall identify $\mathfrak{R}$ with a subring of $\mathfrak{R}_n$ so that in place of aI_n, where $a \in \mathfrak{R}$ and I_n is the unit element of $\mathfrak{R}_n$, we shall merely write a. If $A \in \mathfrak{R}_n$, we shall denote by $\mathfrak{C}[A]$ the subring of $\mathfrak{R}_n$ generated by A together with elements of $\mathfrak{C}$. Thus the elements of $\mathfrak{C}[A]$ are the polynomials in A with coefficients from $\mathfrak{C}$.

Now let $A = (a_{ij})$ be an element of $\mathfrak{R}_n$. If σ is a set of distinct indices in the range $1, 2, \cdots, n$ and f is an element of σ, we define

$$\mathbf{s}_{\sigma,f}(A) = \sum (-1)^r a_{fh_1} a_{h_1h_2} \cdots a_{h_{r-1}h_r} a_{h_rf},$$

summed over all permutations $h_1, h_2, \cdots, h_r$ of elements of σ other than f. By $\mathfrak{R}_n'$ we shall denote the class of all elements A of $\mathfrak{R}_n$ with the following two properties:

I. *If σ is any set of distinct indices in the range* $1, 2, \cdots, n$ *and f and g are arbitrary elements of σ, then*

$$\mathbf{s}_{\sigma,f}(A) = \mathbf{s}_{\sigma,g}(A).$$

II. *For each σ and each element f in σ,* $\mathbf{s}_{\sigma,f}(A) \in \mathfrak{C}$.

Henceforth, if $A \in \mathfrak{R}_n'$, we may merely write $\mathbf{s}_\sigma(A)$ in place of $\mathbf{s}_{\sigma,f}(A)$ and, if there is no question as to what matrix A is under consideration, we shall use $\mathbf{s}_\sigma$ in place of $\mathbf{s}_\sigma(A)$.

If now A is a fixed element of $\mathfrak{R}_n'$, we may define the determinant of A as follows:[3]

$$|A| = \sum \mathbf{s}_{\sigma_1}\mathbf{s}_{\sigma_2} \cdots \mathbf{s}_{\sigma_k},$$

summed on all partitions of $1, 2, \cdots, n$ into disjoint sets σ_i. If $A \in \mathfrak{C}_n$, it may be verified that this definition agrees with the usual one.

Before proceeding, it will be appropriate to investigate the class $\mathfrak{R}_n'$ somewhat more fully. It is to be noted that properties I and II are satisfied by arbitrary elements of $\mathfrak{C}_n$. In fact, if σ consists of two elements, say f and g, property I merely states that $a_{fg}a_{gf} = a_{gf}a_{fg}$, while property II asserts that the product $a_{fg}a_{gf}$ is commutative with all elements of $\mathfrak{R}$. If σ contains more than two elements, the properties I and II are certain symmetry properties which are weaker than full commutativity but are automatically satisfied if commutativity is assumed.

[3] This is Moore's definition. See [6], p. 115.

If $\mathfrak{R}$ is an involutorial ring, then $\mathfrak{R}_n'$ contains all Hermitian matrices of $\mathfrak{R}_n$.[4] It is rather easy to make up individual examples of elements of $\mathfrak{R}_n'$, for suitable choice of $\mathfrak{R}$, but these two classes—the arbitrary matrices over a commutative ring and the Hermitian matrices over an involutorial ring—are the only extensive classes of elements of $\mathfrak{R}_n'$ which have been found, with the exception of such as can be obtained from these by the simple use of direct sums.

We may now state

THEOREM 1. (i) *If* $A \epsilon \mathfrak{R}_n'$ *and* $c \epsilon \mathfrak{C}$, *then* $cA \epsilon \mathfrak{R}_n'$ *and* $|cA| = c^n |A|$; (ii) *if* $cA \epsilon \mathfrak{R}_n'$, $c \epsilon \mathfrak{C}$ *and* c *is not a divisor of zero in* $\mathfrak{R}$, *then* $A \epsilon \mathfrak{R}_n'$; (iii) *if* $A \epsilon \mathfrak{R}_n'$, *then* $\lambda \pm A \epsilon \mathfrak{R}[\lambda]_n'$, *where* λ *is an indeterminate*; (iv) *if* $A \epsilon \mathfrak{R}_n'$, *then any principal minor of* A *of order* s *is an element of* $\mathfrak{R}_s'$.

These statements are almost obvious, except possibly for (ii). If $cA \epsilon \mathfrak{R}_n'$, $c \epsilon \mathfrak{C}$, then

$$\mathbf{s}_{\sigma,f}(cA) = \mathbf{s}_{\sigma,g}(cA) = c^{r+1}\mathbf{s}_{\sigma,f}(A) = c^{r+1}\mathbf{s}_{\sigma,g}(A),$$

where $r + 1$ is the number of elements in σ. From this it follows, under assumption that c is not a divisor of zero in $\mathfrak{R}$, that

$$\mathbf{s}_{\sigma,f}(A) = \mathbf{s}_{\sigma,g}(A) = \mathbf{s}_\sigma(A).$$

Furthermore, since $c^{r+1}\mathbf{s}_\sigma(A) \epsilon \mathfrak{C}$, it follows that for any element x in $\mathfrak{R}$,

$$xc^{r+1}\mathbf{s}_\sigma(A) = c^{r+1}\mathbf{s}_\sigma(A)x,$$

or

$$c^{r+1}[x\mathbf{s}_\sigma(A) - \mathbf{s}_\sigma(A)x] = 0.$$

Hence

$$x\mathbf{s}_\sigma(A) = \mathbf{s}_\sigma(A)x,$$

and thus $\mathbf{s}_\sigma(A) \epsilon \mathfrak{C}$. Thus A satisfies both properties I and II and is therefore an element of $\mathfrak{R}_n'$.

Now a study of Moore's work ([6], pp. 116–124) reveals the fact that, although the taking of conjugates is frequently suggested as a simple way to obtain one part of a theorem from another, this process can be easily avoided throughout and all the theorems there established are true for elements of $\mathfrak{R}_n'$. The only place where the nature of the ring, or the Hermitian character of the matrices, plays any essential part is in the proof of Lemma 16.2, which is precisely what we have assumed in properties I and II. However, before being able to establish for elements of $\mathfrak{R}_n'$ the theorems concerning the determinant of adj A to be found on p. 125 of [6], it is necessary to show that if A is an element of $\mathfrak{R}_n'$, so is adj A. This result we shall establish in the next section, which is the only place in the paper we shall need to make detailed use of Moore's notation.

3. The main theorem about adjoints.

Let σ denote a set of distinct indices in the range $1, 2, \cdots, n$. By $-\sigma$ we mean the elements of the set $1, 2, \cdots, n$

[4] This follows from Moore-Barnard [6], p. 114 and Jacobson [2].

which are not in σ. If f and g are in σ, we denote the set of elements of σ other than f and g by $\sigma - (f, g)$.

If $A \in \mathfrak{R}'_n$, by A^σ we shall indicate the matrix obtained from A by striking out all rows and columns in $-\sigma$. Thus A^σ is a principal minor of A, and its determinant may be denoted by a^σ. If $\sigma = (1, 2, \cdots, n)$, the symbol σ will be omitted, thus $|A| = a$. We define $a^{(0)} = 1$. We may now define adj $A^\sigma = (b^\sigma_{fg})$ as follows:[5]

$$b^\sigma_{ff} = a^{\sigma - f} \qquad (f \text{ in } \sigma),$$

$$b^\sigma_{fg} = \sum_{s=0}^{n_\sigma - 2} \sum (-1)^{s+1} a_{fh_1} a_{h_1h_2} \cdots a_{h_{s-1}h_s} a_{h_sg} a^{\sigma-(f,g,h_1,\cdots,h_s)},$$

where $f \neq g$ and f and g are elements of σ. Here n_σ is the number of elements in σ and the second sum is taken over all permutations $h_1, \cdots, h_s$ of each distinct combination of s elements of $\sigma - (f, g)$. It is understood that, if $s = 0$, we mean simply $a_{fg}a^{\sigma-(f,g)}$. This definition is given here merely for the sake of completeness as we shall not have occasion to make use of it, but shall rather use one of Moore's theorems based on it, which is seen to be true for elements of $\mathfrak{R}'_n$ by the remarks at the end of the preceding section. This theorem will be stated below as Lemma 1. The main purpose of the present section is to prove

THEOREM 2. *If $A \in \mathfrak{R}'_n$, then* adj $A \in \mathfrak{R}'_n$.

The proof will be based on several lemmas which we shall presently establish but first we give a short outline of the proof. Let m be an arbitrary positive integer not exceeding $n - 1$; let $i_1, i_2, \cdots, i_m$ be fixed distinct integers from the set $1, 2, \cdots, n$; and let f, g be also from this set but distinct from $i_1, i_2, \cdots, i_m$. We do not assume that f and g are necessarily distinct. Let us set

$$(1) \qquad S(i_1, i_2, \cdots, i_m; f, g) = \sum b_{fh_1}b_{h_1h_2} \cdots b_{h_{m-1}h_m}b_{h_mg},$$

summed on all permutations $h_1, h_2, \cdots, h_m$ of $i_1, i_2, \cdots, i_m$. In view of the definition of $\mathfrak{R}'_n$ by means of properties I and II, we only need to show that if $A \in \mathfrak{R}'_n$, then $S(i_1, i_2, \cdots, i_m; f, f)$ is symmetric in $i_1, \cdots, i_m, f$ and is also an element of $\mathfrak{C}$. We shall obtain, by induction on m, a formula for $S(i_1, \cdots, i_m; f, g)$ and then show that if f and g are identical, this expression belongs to $\mathfrak{C}$ and is unchanged under any permutation of $i_1, \cdots, i_m, f$.

As indicated above, our first lemma is a restatement of Theorem 16.7 of [6] as follows:

LEMMA 1. *If $A \in \mathfrak{R}'_n$, σ is a set of distinct indices in the range* $1, 2, \cdots, n$ *and f, g are in σ, then*

$$a^\sigma b_{fg} = ab^\sigma_{fg} + \sum_h^{-\sigma} b^{(\sigma,h)}_{fh} b_{hg},$$

it being understood that this sum is extended over all elements h of $-\sigma$.

[5] This is the definition to be found in [6], p. 119, with a slight change in notation to avoid introducing part of the notation used there.

As an illustration of the application of this lemma, suppose $\sigma = -(i_1)$. Then, in the notation introduced above, we find

$$S(i_1 ; f, g) = a^{-(i_1)} b_{fg} - a b_{fg}^{-(i_1)}, \tag{2}$$

which becomes, for $g = f$,

$$S(i_1 ; f, f) = a^{-(i_1)} a^{-(f)} - a a^{-(i_1, f)}.$$

This is clearly in $\mathfrak{C}$ as each term on the right is a product of determinants of principal minors of A. Also, since these determinants are in $\mathfrak{C}$, it is clear that $S(i_1 ; f, f)$ is unchanged by the interchange of i_1 and f. It is this kind of calculation that we need to generalize, and the main part of the work consists in generalizing formula (2). First we need some additional notation.

Let m be a fixed positive integer not exceeding $n - 1$ and $i_1, i_2, \cdots, i_m$ a fixed set of distinct integers from the set $1, 2, \cdots, n$. Let f and g, not necessarily distinct, be integers from the set $1, 2, \cdots, n$ but different from $i_1, i_2, \cdots, i_m$. Let $k_1, \cdots, k_r$ be *positive* integers, not necessarily distinct, such that $k_1 + k_2 + \cdots + k_r \leqq m$, and set $m - (k_1 + \cdots + k_r) = k \geqq 0$. The case in which $r = 0$ is not excluded, and we mean in this case that $k = m$ and there are no elements in the set $k_1, \cdots, k_r$. Let us set

$$\sigma_1 = (i_1, \cdots, i_{k_1}),\ \sigma_2 = (i_{k_1+1}, \cdots, i_{k_1+k_2}), \qquad \cdots,$$

$$\sigma_r = (i_{k_1+\cdots+k_{r-1}+1}, \cdots, i_{k_1+\cdots+k_r}), \qquad \sigma_{r+1} = (i_{k_1+\cdots+k_r+1}, \cdots, i_m).$$

Thus there are k_1 elements in $\sigma_1, \cdots, k_r$ elements in σ_r and k elements in σ_{r+1}. We now introduce the notation

$$[k_1, k_2, \cdots, k_r ; k] = a^{m-r} \sum a^{-\sigma_1} a^{-\sigma_2} \cdots a^{-\sigma_r} b_{fg}^{-\sigma_{r+1}}, \tag{3}$$

where the sum is to be taken over all permutations of $i_1, i_2, \cdots, i_m$ which take the expression

$$x_{i_1} x_{i_2} \cdots x_{i_{k_1}} + \cdots + x_{i_{k_1+\cdots+k_{r-1}+1}} \cdots x_{i_{k_1+\cdots+k_r}} + y_{i_{k_1+\cdots+k_r+1}} \cdots y_{i_m} \tag{4}$$

into distinct such expressions, it being understood that $x_{i_1}, \cdots, x_{i_m}, y_{i_1}, \cdots, y_{i_m}$ are distinct commutative indeterminates. We may remark that (4) is unchanged under permutation of the elements in any single product and also under interchange of two products of x's which have the same number of letters. Also each factor of every term on the right side of (3), with the exception of factors of the form $b_{fg}^{-\sigma_{r+1}}$, are in $\mathfrak{C}$ and can therefore be arranged at pleasure. Hence $[k_1, \cdots, k_r ; k]$ is unchanged under any permutation of $k_1, k_2, \cdots, k_r$.

The main part of the proof of Theorem 2 is contained in the proof of

LEMMA 2. *We have*

$$S(i_1, \cdots, i_m ; f, g) = \sum r!(-1)^{m-r} [k_1, k_2, \cdots, k_r ; k], \tag{5}$$

summed over all different sets[6] $k_1, \cdots, k_r$ *of positive integers, repetitions being*

[6] Two sets are *different* if the elements of one can not be obtained by permuting the elements of the other.

allowed, such that $k_1 + \cdots + k_r \leqq m$. *As above,* $k = m - (k_1 + \cdots + k_r)$, $r \geqq 0$.

The proof is by induction on m. If $m = 1$, (5) reduces to (2), which is true by Lemma 1. Accordingly, we assume (5) for a fixed $m < n - 1$ and shall prove it for $m + 1$. An examination of (1) shows that we can pass from $S(i_1, \cdots, i_m; f, g)$ to $S(i_1, \cdots, i_{m+1}; f, g)$ by the following *induction operation.* Replace g by i_{m+1}, multiply by $b_{i_{m+1}g}$ on the right and then in the resulting expression perform in turn the transpositions $(i_{m+1}, i_1), \cdots, (i_{m+1}, i_m)$ and add. We need to investigate closely the result of this operation on the right side of (5). To this end, we consider first a single term $[k_1, \cdots, k_r; k]$ occurring on the right side of (5) and shall prove

LEMMA 3. *By the above defined induction operation,* $[k_1, \cdots, k_r; k]$ *goes over into*

$$-[k_1, \cdots, k_r; k + 1] + \alpha[k_1, \cdots, k_r, k + 1; 0],$$

where these symbols are defined as in (3) *with* m *replaced by* $m + 1$ *and* α *is one more than the number of the integers* $k_1, \cdots, k_r$ *which are equal to* $k + 1$.

If we perform the induction operation on $[k_1, \cdots, k_r; k]$, we get

$$a^{m-r} \sum\sum a^{-\sigma_1} a^{-\sigma_2} \cdots a^{-\sigma_r} b_{f i_{m+1}}^{-\sigma_{r+1}} b_{i_{m+1}g}, \tag{6}$$

where the inner sum is precisely the sum appearing in (3) while the outer one means a sum over the successive transpositions $(i_{m+1}, i_1), \cdots, (i_{m+1}, i_m)$, (i_{m+1}, i_{m+1}). Evidently (6) is symmetric in $i_1, \cdots, i_{m+1}$. Let us find the coefficient of $a^{m-r}a^{-\sigma_1} \cdots a^{-\sigma_r}$ in (6). A similar result will follow after any permutation of $i_1, i_2, \cdots, i_{m+1}$. Clearly,

$$a^{m-r}a^{-\sigma_1}a^{-\sigma_2} \cdots a^{-\sigma_r} \sum b_{f i_{m+1}}^{-\sigma_{r+1}} b_{i_{m+1}g}, \tag{7}$$

where this sum is taken over the successive interchanges of i_{m+1} with elements of σ_{r+1}, appears in (6). But, by Lemma 1 with $\sigma = -(\sigma_{r+1}, i_{m+1})$, the expression (7) is equal to

$$a^{m-r}a^{-\sigma_1}a^{-\sigma_2} \cdots a^{-\sigma_r}[a^{-(\sigma_{r+1}, i_{m+1})} b_{fg} - a b_{fg}^{-(\sigma_{r+1}, i_{m+1})}]. \tag{8}$$

Thus (8) is a sum of two terms, one from $[k_1, \cdots, k_r, k + 1; 0]$ and the other from $-[k_1, \cdots, k_r; k + 1]$. It will be observed that we can obtain in this way from (6) each term of $-[k_1, \cdots, k_r; k + 1]$ once and only once. However, each term of $[k_1, \cdots, k_r, k + 1; 0]$ is obtained once more than the number of times $k + 1$ appears among the set $k_1, k_2, \cdots, k_r$. For example, if $k_1 = k + 1$, there will be another expression similar to (8) with σ_1 interchanged with (σ_{r+1}, i_{m+1}). The first term of this expression will actually be equal to the first term of (8), but the second term will be a term of $-[k_1, \cdots, k_r; k + 1]$ different from the one appearing in (8). These considerations establish Lemma 3 and we proceed to complete the proof of Lemma 2.

In view of our hypothesis that (5) is true, and Lemma 3, we have

$$(9)\quad S(i_1, \cdots, i_{m+1}; f, g) = \sum r!(-1)^{m-r}\{\alpha[k_1, \cdots, k_r, k+1; 0] - [k_1, \cdots, k_r; k+1]\},$$

the sum being precisely the sum in (5). We wish to prove

$$(10)\qquad S(i_1, \cdots, i_{m+1}; f, g) = \sum s!(-1)^{m+1-s}[l_1, \cdots, l_s; l],$$

summed over all sets $l_1, l_2, \cdots, l_s$ of positive integers such that $l_1 + l_2 + \cdots + l_s \leqq m + 1$, where $s \geqq 0$ and $l = m + 1 - (l_1 + \cdots + l_s)$. To this end we choose fixed integers $l_1, l_2, \cdots, l_s, l$ satisfying these conditions and seek where $[l_1, \cdots, l_s; l]$ occurs in (9). If $l \neq 0$, it appears once and only once, namely, in the term with $k_i = l_i$ $(i = 1, 2, \cdots, s)$, $k = l - 1$, and the coefficient is $s!(-1)^{m+1-s}$ as required in (10). Suppose now that $l = 0$, and let us assume that $l_{t_1}, \cdots, l_{t_j}$ are the distinct integers among $l_1, l_2, \cdots, l_s$ and that l_{t_1} appears p_1 times, l_{t_2} appears p_2 times, and so on. Thus $p_1 + \cdots + p_{t_j} = s$. Then $[l_1, \cdots, l_s; 0]$ appears in (9) with $k_i = l_i$ $(i = 2, 3, \cdots, s)$, $k = l_1 - 1$, with a coefficient $p_1(s-1)!(-1)^{m-s+1}$. Also this expression appears in (9) when $k_i = l_i$ $(i \neq t_2)$, $k = l_{t_2} - 1$, with a coefficient $p_2(s-1)!(-1)^{m-s+1}$, and so on. Adding all these coefficients, we see that $[l_1, \cdots, l_s; 0]$ appears in (9) with a coefficient $s!(-1)^{m+1-s}$, and this is the desired coefficient. The proof is therefore completed.

We may now complete the proof of Theorem 2. If, in formula (5), we replace g by f and expand each expression of the type $[k_1, \cdots, k_r; k]$ into a sum of terms as given in (3), we see that, since by definition $b_{ff}^{-\sigma_r+1} = a^{-(\sigma_{r+1}, f)}$, every term is a product of determinants of principal minors of A. Since these are all in $\mathfrak{C}$, it follows that $S(i_1, \cdots, i_m; f, f) \in \mathfrak{C}$. There remains only to show that $S(i_1, \cdots, i_m; f, f)$ is symmetric in $i_1, \cdots, i_m, f$. It is clearly symmetric in $i_1, \cdots, i_m$ and we therefore only need to show that it is unchanged under the interchange of f and i_1. Consider a single term in the expansion of the right side of (5) with $g = f$. If, for example, $m = 5$ and i_1 appears with f as in

$$a^{-(i_2, i_4)}a^{-(i_3)}a^{-(i_1, i_5, f)},$$

this term is clearly unchanged under the interchange of i_1 and f. Corresponding to any term in which i_1 and f do not appear together, say

$$a^{-(i_1, i_2)}a^{-(i_3)}a^{-(i_4, i_5, f)}$$

which comes from [1, 2; 2], there is another term

$$a^{-(i_3)}a^{-(i_4, i_5, i_1)}a^{-(i_2, f)}$$

with i_1 and f interchanged. This term comes from [1, 3; 1] and has the same coefficient in (5) as the preceding term. Their sum is obviously unchanged under the interchange of i_1 and f. In general, one can thus pair all terms occurring in the right of (5) in which i_1 and f do not appear together in such a

way that the interchange of i_1 and f leaves the sum of each pair unchanged. This completes the proof of Theorem 2.

We may remark that, in the notation of §2, our proof really shows that, if $A \epsilon \mathfrak{R}_n'$, then $\mathbf{s}_\sigma$ (adj A) is expressible as a polynomial, with integral coefficients, in the determinants of the principal minors of A. In view of our definition of determinant, this means finally that $\mathbf{s}_\sigma$ (adj A) is expressible as a polynomial, with integral coefficients, in the different $\mathbf{s}_{\sigma'}(A)$.

4. **Further properties of** adj A. In view of Theorem 2, it may now be verified that the theorems in [6], pp. 125–127, having to do with the determinant of adj A are valid for elements of $\mathfrak{R}_n'$. We shall explicitly mention only two results which will be used in the sequel, and which are well known for matrices over a commutative ring. These are as follows, where it is assumed that $A \epsilon \mathfrak{R}_n'$:

$$A \text{ adj } A = (\text{adj } A)A = |A|, \tag{11}$$

and

$$|\text{adj } A| = |A|^{n-1}. \tag{12}$$

As a matter of fact, the first of these can be established without use of Theorem 2, while the second naturally requires the theorem.

It is now easy to prove

THEOREM 3. *If $A \epsilon \mathfrak{R}_n'$ and A is a divisor of zero in $\mathfrak{R}_n$, then $|A|$ is a divisor of zero in $\mathfrak{R}$.*

For if $AX = 0$, $X \neq 0$, multiplication by adj A on the left yields at once $|A|X = 0$.

If now $A \epsilon \mathfrak{R}_n'$ and λ is an indeterminate, we shall frequently make use of the *characteristic polynomial* of A,

$$f(\lambda) = |\lambda - A| = \lambda^n + a_1\lambda^{n-1} + \cdots + a_n. \tag{13}$$

It is to be noted that the coefficients in $f(\lambda)$ are from $\mathfrak{C}$ and, in fact, a_i is, except possibly for sign, the sum of the determinants of the principal minors of A of order i. It is easy to show that $f(A) = 0$, which will be a special instance of a more general theorem to be established below, and we shall for the present assume this result. We shall now prove

THEOREM 4. *If $A \epsilon \mathfrak{R}_n'$ and the characteristic polynomial of A is given by formula* (13), *then*

$$\text{adj } A = (-1)^{n+1}(A^{n-1} + a_1A^{n-2} + \cdots + a_{n-1}). \tag{14}$$

Adjoin an indeterminate μ to $\mathfrak{R}$, getting the ring $\mathfrak{R}[\mu]$. From (11), applied to $\mu + A$ which, by Theorem 1, is in $\mathfrak{R}[\mu]_n'$, we get

$$(\mu + A) \text{ adj } (\mu + A) = |\mu + A|. \tag{15}$$

Now $\mu + A$ satisfies its characteristic equation, say

$$(\mu + A)^n + b_1(\mu + A)^{n-1} + \cdots + b_n = 0,$$

where $b_n = (-1)^n \mid \mu + A \mid$. This may be rewritten in the form

$$(\mu + A)[(\mu + A)^{n-1} + \cdots + b_{n-1}] = (-1)^{n+1} \mid \mu + A \mid. \tag{16}$$

From (15) and (16) it therefore follows that

$$(\mu + A)\{(-1)^{n+1} \operatorname{adj} (\mu + A) - [(\mu + A)^{n-1} + \cdots + b_{n-1}]\} = 0.$$

Now $\mid \mu + A \mid = \mu^n + \cdots$ is clearly not a divisor of zero in $\mathfrak{R}[\mu]$, and thus, by Theorem 3, $\mu + A$ is not a divisor of zero in $\mathfrak{R}[\mu]_n$. Hence

$$\operatorname{adj} (\mu + A) = (-1)^{n+1}[(\mu + A)^{n-1} + \cdots + b_{n-1}], \tag{17}$$

and the desired result follows by equating the terms on both sides of (17) which are independent of μ or, what amounts to the same thing, formally replacing μ by 0.

From relation (11) and Theorem 2, it is obvious that if $A \in \mathfrak{R}'_n$ and $\mid A \mid$ has an inverse in $\mathfrak{R}$ (actually in $\mathfrak{C}$), then A has an inverse in $\mathfrak{R}'_n$, namely $\mid A \mid^{-1} \operatorname{adj} A$. We may also prove the following partial converse.

THEOREM 5. *If $A \in \mathfrak{R}'_n$ has an inverse in $\mathfrak{R}_n$ and $\mid A \mid$ is not a divisor of zero in $\mathfrak{R}$, then $\mid A \mid$ has an inverse in $\mathfrak{R}$.*

If $AX = 1$, and we multiply on the left by adj A, we get

$$\mid A \mid X = \operatorname{adj} A. \tag{18}$$

Thus, by Theorem 2, $\mid A \mid X \in \mathfrak{R}'_n$ and, since $\mid A \mid$ is not a divisor of zero in $\mathfrak{R}$, this implies by (ii) of Theorem 1 that $X \in \mathfrak{R}'_n$. Taking determinants of both sides of (18), we have

$$\mid A \mid^n \mid X \mid = \mid A \mid^{n-1}.$$

This, however, implies that $\mid A \mid \cdot \mid X \mid = 1$, and the theorem is established.

5. **Minimum ideal and related topics.** We now pass to a generalization of the familiar notion of minimum function of a matrix with coefficients in a field.

If $A \in \mathfrak{R}_n$ and $g(\lambda) = a_0 + a_1\lambda + \cdots + a_m\lambda^m$ is an element of $\mathfrak{R}[\lambda]$, we define

$$g_r(A) = a_0 + a_1A + \cdots + a_mA^m$$

and

$$g_l(A) = a_0 + Aa_1 + \cdots + A^m a_m .$$

It is easy to see that the set of all elements $g(\lambda)$ of $\mathfrak{R}[\lambda]$ such that $g_r(A) = 0$ is a left ideal $\mathfrak{m}_l$ in $\mathfrak{R}[\lambda]$. Similarly, we may denote by $\mathfrak{m}_r$ the right ideal in $\mathfrak{R}[\lambda]$ of all those elements $g(\lambda)$ such that $g_l(A) = 0$. The totality of all elements $h(\lambda)$ of $\mathfrak{C}$ such that $h(A) = 0$ is an ideal $\mathfrak{m}$ in $\mathfrak{C}[\lambda]$, which we shall henceforth

call the *minimum ideal* of A since it plays a rôle similar to the ordinary minimum function of a matrix if the elements are from a field. Clearly $\mathfrak{m} = \mathfrak{m}_l \cap \mathfrak{C}[\lambda] = \mathfrak{m}_r \cap \mathfrak{C}[\lambda]$. It is also clear that $\mathfrak{C}[A] \cong \mathfrak{C}[\lambda]/\mathfrak{m}$, so that the ring $\mathfrak{C}[A]$ is determined when the ideal $\mathfrak{m}$ is characterized. In general, the determination of $\mathfrak{m}_l$, $\mathfrak{m}_r$ and $\mathfrak{m}$ may be quite difficult, but if $A \in \mathfrak{R}'_n$ we shall now show how these ideals may be determined.

Let A be a fixed element of $\mathfrak{R}'_n$, with characteristic polynomial $f(\lambda) = |\lambda - A|$, and let us set adj $(\lambda - A) = (h_{ij}(\lambda))$. We now prove[7]

THEOREM 6. *If $A \in \mathfrak{R}'_n$, then $g_r(A) = 0$ if and only if*

$$g(\lambda)h_{ij}(\lambda) \equiv 0 \qquad (f(\lambda)) \qquad (i, j = 1, 2, \cdots, n). \tag{19}$$

Similarly, $g_l(A) = 0$ if and only if

$$h_{ij}(\lambda)g(\lambda) \equiv 0 \qquad (f(\lambda)) \quad (i, j = 1, 2, \cdots, n). \tag{20}$$

We shall prove the first part of the theorem. If we assume relation (19), we have

$$g(\lambda) \text{ adj } (\lambda - A) = Bf(\lambda),$$

where $B \in \mathfrak{R}[\lambda]_n$. Multiply this equation by $\lambda - A$ on the right, thus getting

$$g(\lambda)f(\lambda) = Bf(\lambda)(\lambda - A) = B(\lambda - A)f(\lambda).$$

Now $f(\lambda)$ has leading coefficient 1 and is therefore not a divisor of zero. Hence

$$g(\lambda) = B(\lambda - A), \tag{21}$$

and the factor theorem (cf. [1], p. 26) shows at once that $g_r(A) = 0$.

Conversely, if $g_r(A) = 0$, the factor theorem states the existence of a relation (21) and multiplication by adj $(\lambda - A)$ on the right yields

$$g(\lambda) \text{ adj } (\lambda - A) = Bf(\lambda), \tag{22}$$

which is equivalent to (19). The second part of the theorem can be established by a similar argument.

We may easily establish the following

COROLLARY. *If $A \in \mathfrak{R}'_n$ and $g(\lambda) \equiv 0$ $(\mathfrak{m})$, then*

$$[g(\lambda)]^n \equiv 0 \qquad (f(\lambda)),$$

and thus $\mathfrak{m}$ and $f(\lambda)$ have the same prime ideal divisors in $\mathfrak{C}[\lambda]$.

If $g(\lambda) \equiv 0$ $(\mathfrak{m})$, then by Theorem 2 the left side of equation (22) is in $\mathfrak{R}[\lambda]'_n$, and the same is therefore true of the right side. Since $f(\lambda)$ is not a divisor of zero in $\mathfrak{R}[\lambda]$, this implies, by (ii) of Theorem 1, that $B \in \mathfrak{R}[\lambda]'_n$. We may therefore take determinants of both sides of (22), getting

$$[g(\lambda)]^n[f(\lambda)]^{n-1} = |B| [f(\lambda)]^n,$$

from which it follows that

$$[g(\lambda)]^n = |B| f(\lambda).$$

[7] This is essentially the proof for the commutative case to be found in [5] and it is included here for the sake of completeness.

6. **Generalization of Ostrowski's theorem.** Recently, Ostrowski [7] has extended a theorem due to Phillips [8] and thus obtained for the case of several matrices with coefficients in a field an analogue of the notion of minimum function. We shall now show how this result can be extended to the general case under discussion in this paper. The case of an arbitrary commutative ring has already been treated in [4], and we shall present in detail only those parts which have to be essentially modified in the non-commutative case.

Let $x_1, \cdots, x_m$ be commutative indeterminates and let us denote by $\mathfrak{S}$ the ring $\mathfrak{R}[x_1, \cdots, x_m]$. Throughout this section, *we assume that* A_k $(k = 1, 2, \cdots, m)$ *are fixed elements of* $\mathfrak{R}_n$ *such that the matrix*

$$x_1A_1 + x_2A_2 + \cdots + x_mA_m \tag{23}$$

is an element of $\mathfrak{S}_n'$. This implies, in particular, that each A_k is in $\mathfrak{R}_n'$. The above stated condition will certainly be satisfied if $\mathfrak{R}$ is a commutative ring and the A_k are arbitrary elements of $\mathfrak{R}_n$ or if $\mathfrak{R}$ is an involutorial ring and the A_k are Hermitian.

Let us set

$$| x_1A_1 + \cdots + x_mA_m | = F(x_1, \cdots, x_m),$$

which is an element in the center of $\mathfrak{S}$. Further, let

$$\operatorname{adj}\,(x_1A_1 + \cdots + x_mA_m) = (F_{ij}(x_1, \cdots, x_m)).$$

Denote by $\mathfrak{n}_l$ the left ideal in $\mathfrak{S}$ of those elements $f(x_1, \cdots, x_m)$ such that

$$f(x_1, \cdots, x_m)F_{ij}(x_1, \cdots, x_m) \equiv 0 \quad (F(x_1, \cdots, x_m)) \quad (i, j = 1, 2, \cdots, n). \tag{24}$$

Similarly, let $\mathfrak{n}_r$ be the right ideal consisting of those elements $f(x_1, \cdots, x_m)$ of $\mathfrak{S}$ such that

$$F_{ij}(x_1, \cdots, x_m)f(x_1, \cdots, x_m) \equiv 0 \quad (F(x_1, \cdots, x_m)) \quad (i, j = 1, 2, \cdots, n). \tag{25}$$

We may now prove

THEOREM 7. *Let* A_k $(k = 1, 2, \cdots, m)$ *be elements of* $\mathfrak{R}_n$ *such that the matrix* (23) *is an element of* $\mathfrak{S}_n'$ *and such that* $F(x_1, \cdots, x_m)$ *is not a divisor of zero in* $\mathfrak{S}$. *Let* $B_1, B_2, \cdots, B_m$ *be elements of* $\mathfrak{R}_n$, *commutative in pairs, such that*

$$A_1B_1 + A_2B_2 + \cdots + A_mB_m = 0. \tag{26}$$

Then if $f(x_1, \cdots, x_m) \equiv 0$ $(\mathfrak{n}_l)$, *it follows that* $f_r(B_1, \cdots, B_m) = 0$.[8] *Similarly, if*

$$B_1A_1 + B_2A_2 + \cdots + B_mA_m = 0 \tag{27}$$

and $f(x_1, \cdots, x_m) \equiv 0$ $(\mathfrak{n}_r)$, *then* $f_l(B_1, \cdots, B_m) = 0$.

We may remark first that although it is not a necessary condition, $F(x_1, \cdots, x_m)$ will not be a divisor of zero in $\mathfrak{S}$ if, for any k, $| A_k |$ is not a

[8] By this notation, we mean that in $f(x_1, \cdots, x_m)$ we replace x_i by B_i $(i = 1, 2, \cdots, m)$ and write the power products of the B's on the *right* in each term.

divisor of zero in $\mathfrak{R}$. This is evident since the term in $F(x_1, \cdots, x_m)$ of highest degree in x_k is precisely $|A_k| x_k^n$.

We shall prove the first part of the theorem only, as the second part can be proved similarly. The proof is an adaptation of the proof given in [4] for the commutative case.

Let us set $A_k = (a_{ij}^{(k)})$. By hypothesis that $f(x_1, \cdots, x_m) \equiv 0 \ (\mathfrak{n}_l)$, it follows that

$$f(x_1, \cdots, x_m)F_{ij}(x_1, \cdots, x_m) = h_{ij}(x_1, \cdots, x_m)F(x_1, \cdots, x_m) \qquad (i, j = 1, 2, \cdots, n), \tag{28}$$

where the $h_{ij}(x_1, \cdots, x_m)$ are elements of $\mathfrak{S}$. Furthermore, by formula (11) we have

$$\sum_{l=1}^{n} F_{il}(x_1, \cdots, x_m) \sum_{k=1}^{m} x_k a_{lj}^{(k)} = \delta_{ij} F(x_1, \cdots, x_m).$$

If we multiply this by $f(x_1, \cdots, x_m)$ on the left, make use of equation (28) and cancel $F(x_1, \cdots, x_m)$ from both sides since it is not a divisor of zero in $\mathfrak{S}$, we get

$$\sum_{l=1}^{n} h_{il}(x_1, \cdots, x_m) \sum_{k=1}^{m} x_k a_{lj}^{(k)} = \delta_{ij} f(x_1, \cdots, x_m). \tag{29}$$

Now let us set

$$C_{lj} = \sum_{k=1}^{m} a_{lj}^{(k)} B_k,$$

and

$$h_{il}(x_1, \cdots, x_m) = \sum_{\alpha} h_{il}^{(\alpha)} x_{il}^{(\alpha)},$$

where $x_{il}^{(\alpha)}$ indicates the different power products of the x's which appear in $h_{il}(x_1, \cdots, x_m)$. This notation merely serves the purpose of enabling us to separate the coefficients in $h_{il}(x_1, \cdots, x_m)$ from the power products of the x's. Now, in (29), the x's are commutative with everything and if we write them on the right in every term on both sides of (29), the equation (29) then says merely that the coefficient of any power product of the x's is the same on both sides. The equation therefore remains true if we replace x_k by B_k $(k = 1, 2, \cdots, m)$, thus getting

$$\sum_{l=1}^{n} \sum_{\alpha} h_{il}^{(\alpha)} C_{lj} B_{il}^{(\alpha)} = \delta_{ij} f_r(B_1, \cdots, B_m). \tag{30}$$

It is naturally understood that $B_{il}^{(\alpha)}$ represents the same power product of the B's that $x_{il}^{(\alpha)}$ is of the x's.

Now if E_{ij} is the matrix with 1 at the intersection of the i-th row and j-th column and zeros elsewhere, we may rewrite (26) as follows:

$$0 = \sum_{k=1}^{m} \sum_{i,j=1}^{n} a_{ij}^{(k)} E_{ij} B_k = \sum_{i,j=1}^{n} E_{ij} \sum_{k=1}^{m} a_{ij}^{(k)} B_k = \sum_{i,j=1}^{n} E_{ij} C_{ij}.$$

Multiplication by E_{1l} on the left yields

$$0 = \sum_{i,j=1}^{n} E_{1l} E_{ij} C_{ij} = \sum_{j=1}^{n} E_{1j} C_{lj}.$$

Now multiply this equation by $h_{il}^{(\alpha)}$ on the left, $B_{il}^{(\alpha)}$ on the right, and sum on α and l. The matrix E_{1j} is commutative with $h_{il}^{(\alpha)}$, and we therefore get, by making use of (30),

$$0 = \sum_{j=1}^{n} E_{1j} \sum_{\alpha,l} h_{il}^{(\alpha)} C_{lj} B_{il}^{(\alpha)} = \sum_{j=1}^{n} E_{1j} \delta_{ij} f_r(B_1, \cdots, B_m)$$
$$= E_{1i} f_r(B_1, \cdots, B_m).$$

If we multiply this by E_{i1} on the left and sum on i, it follows that $f_r(B_1, \cdots, B_m) = 0$, which is the desired result.

It is possible to prove a partial converse of the preceding theorem by making only trivial changes in the proof for the commutative case (cf. [4], p. 494). Accordingly, we shall merely state this result without proof as

THEOREM 8. *We assume that $A_1 = 1$, and also that $\mathfrak{R}$ has the property that, if $g(\lambda) \in \mathfrak{R}[\lambda]$ such that $g(a) = 0$ for every element a of $\mathfrak{C}$, then $g(\lambda) = 0$. If $f(x_1, \cdots, x_m)$ is an element of $\mathfrak{S}$ such that $f_r(B_1, \cdots, B_m) = 0$ for every choice of matrices $B_1, \cdots, B_m$ which are commutative and satisfy (26), then $f(x_1, \cdots, x_m) \equiv 0$ $(\mathfrak{n}_l)$.*

A similar result holds if r and l are interchanged and relation (26) replaced by (27). We may remark also that considerations similar to those used in the proof of the corollary to Theorem 6 show at once that, if $f(x_1, \cdots, x_m)$ is in $\mathfrak{C}[x_1, \cdots, x_m]$ and is an element of $\mathfrak{n}_r$ (or $\mathfrak{n}_l$), then

$$[f(x_1, \cdots, x_m)]^n \equiv 0 \qquad (F(x_1, \cdots, x_m)).$$

7. **Quaternion rings. Definition of class $\mathcal{K}$.** There are a number of theorems which are known to be true for matrices over a commutative ring and which we are unable to establish for arbitrary elements of $\mathfrak{R}_n'$. We proceed to restrict further the class of matrices considered, and to this end we pause to introduce the concept of a *quaternion ring*.

Let $\mathfrak{C}$ be a commutative ring with unit element 1, in which the equation $2x = 1$ has a unique solution $x = \frac{1}{2}$. We consider the linear form module of expressions of the form

$$(31) \qquad a = c_0 w_0 + c_1 w_1 + c_2 w_2 + c_3 w_3 \qquad (c_i \text{ in } \mathfrak{C}),$$

where w_0, w_1, w_2, w_3 are linearly independent over $\mathfrak{C}$ and are commutative with elements of $\mathfrak{C}$. We assume for the w_i the following multiplication table:

$$w_0 w_i = w_i w_0 = w_i \quad (i = 0, 1, 2, 3), \quad w_1^2 = p, \quad w_2^2 = q, \quad w_3^2 = -pq,$$
$$w_1 w_2 = -w_2 w_1 = w_3, \quad w_1 w_3 = -w_3 w_1 = p w_2, \quad w_2 w_3 = -w_3 w_2 = -q w_1,$$

where p and q are elements of $\mathfrak{C}$. We assume further that if x is an element of $\mathfrak{C}$ such that $px = qx = 0$, then $x = 0$; in other words, p and q are not simultaneously annihilated by any non-zero element of $\mathfrak{C}$. Under these hypotheses, the totality of elements of form (31) is a ring $\mathfrak{R}$ which we call a *quaternion ring over* $\mathfrak{C}$ or simply a *quaternion ring*. Clearly $\mathfrak{R}$ contains a subring of all elements of form c_0w_0, which is isomorphic to $\mathfrak{C}$ and which we identify with $\mathfrak{C}$. Thus, henceforth we shall write 1 in place of w_0.

It is easy to show that $\mathfrak{C}$ is the center of $\mathfrak{R}$. For if a, defined by (31), is in the center of $\mathfrak{R}$, then $aw_1 = w_1a$ implies that $a_2 = 0$ and $pa_3 = 0$, while $aw_2 = w_2a$ implies that $a_1 = 0$ and $qa_3 = 0$. By our assumption concerning p and q, it follows that $a_3 = 0$ and hence $a = c_0 \in \mathfrak{C}$. Clearly elements of $\mathfrak{C}$ are in the center of $\mathfrak{R}$ and therefore $\mathfrak{C}$ coincides with the center.

We may naturally define conjugates in the usual way. If a is given by (31), then $a \to \bar{a}$, where

$$\bar{a} = c_0 - c_1w_1 - c_2w_2 - c_3w_3$$

defines an involution in $\mathfrak{R}$, and further the symmetric elements are precisely the elements of $\mathfrak{C}$. Thus $\mathfrak{R}$ is an involutorial ring as defined in the introduction. Obviously, generalized quaternion algebras over a field of characteristic other than 2 are instances of quaternion rings. We may also remark that the complete matric ring $\mathfrak{C}_2$ is a quaternion ring over $\mathfrak{C}$ as can be seen by setting $p = 1$, $q = -1$. Then if E_{ij} denotes the matric units used in the preceding section, we have $E_{11} = \frac{1}{2}(1 + w_1)$, $E_{12} = \frac{1}{2}(w_3 + w_2)$, $E_{21} = \frac{1}{2}(w_3 - w_2)$, $E_{22} = \frac{1}{2}(1 - w_1)$. Thus if

$$a = \begin{pmatrix} f & g \\ h & i \end{pmatrix},$$

it follows that

$$\bar{a} = \begin{pmatrix} i & -g \\ -h & f \end{pmatrix}.$$

This example of $\mathfrak{C}_2$, as an involutorial ring, for the case in which $\mathfrak{C}$ is a field not of characteristic 2, has been used by Jacobson in [2].

Now if A is a Hermitian matrix over the quaternion ring $\mathfrak{R}$ and $g(\lambda) \in \mathfrak{C}[\lambda]$, then $g(A)$ is also Hermitian. Further, if A and B are Hermitian, and $AB = BA$, then also AB is Hermitian. Now it will be found that certain theorems about Hermitian matrices over a quaternion ring coincide in statement with known theorems about arbitrary matrices over a commutative ring. At least the statements of the theorems, and in some cases the proofs also, can be unified by the following notation. We shall henceforth restrict ourselves to the following two situations.

Case 1. $\mathfrak{R}$ is a quaternion ring with center $\mathfrak{C}$. We shall denote by $\mathcal{K}$ the class of all Hermitian matrices of $\mathfrak{R}_n$.

Case 2. $\mathfrak{R}$ is an arbitrary commutative ring with unit element and center $\mathfrak{C} = \mathfrak{R}$. In this case, we let $\mathcal{K} = \mathfrak{R}_n = \mathfrak{C}_n$.

In either case, if $A \epsilon \mathcal{K}$ and $g(\lambda) \epsilon \mathfrak{C}[\lambda]$, then $g(A) \epsilon \mathcal{K}$. Also, if $A \epsilon \mathcal{K}$, then $A \epsilon \mathfrak{R}_n'$ and we have at our disposal all theorems proved for elements of $\mathfrak{R}_n'$.

8. **Some miscellaneous properties of elements of** $\mathcal{K}$. We begin with a theorem which is trivial in Case 2 and which, as a result of Moore's work ([6], p. 147), is also known to be true for Hermitian matrices over a "number system of type B". We shall, in fact, base our proof of the theorem only on the fact that it is true for Hermitian matrices over real quaternions. The result in question is the following

THEOREM 9. *If* $A \epsilon \mathcal{K}$, $g_1(\lambda) \epsilon \mathfrak{C}[\lambda]$, $g_2(\lambda) \epsilon \mathfrak{C}[\lambda]$, *then*

$$| g_1(A)g_2(A) | = | g_1(A) | \cdot | g_2(A) |.$$

We need to establish this result only in the first case. Accordingly, we assume that $A = (a_{ij})$ is a Hermitian matrix over the quaternion ring $\mathfrak{R}$ with center $\mathfrak{C}$. Let

$$g_1(\lambda) = a_0 + a_1\lambda + \cdots + a_k\lambda^k$$

and

$$g_2(\lambda) = b_0 + b_1\lambda + \cdots + b_l\lambda^l.$$

We now need to establish some identities which are useful in establishing the theorem.

Let $\mathfrak{T}$ denote the ring of rationals, and let α, β, x_{ii} $(i = 1, 2, \cdots, n)$, $x_{ij}^{(t)}$ $(t = 0, 1, 2, 3; i, j = 1, 2, \cdots, n; i < j)$, y_i $(i = 0, 1, \cdots, k)$ and z_i $(i = 0, 1, \cdots, l)$ be indeterminates, and let $\mathfrak{T}'$ be the ring obtained from $\mathfrak{T}$ by adjoining all these indeterminates. Further, let $\mathfrak{R}'$ be the quaternion ring over $\mathfrak{T}'$ defined as in the preceding section with α and β taking the place of p and q in the multiplication table. We now introduce a matrix $B = (b_{ij})$, where $b_{ii} = x_{ii}$ $(i = 1, 2, \cdots, n)$, $b_{ij} = x_{ij}^{(0)} + x_{ij}^{(1)}w_1 + x_{ij}^{(2)}w_2 + x_{ij}^{(3)}w_3$ $(i, j = 1, 2, \cdots, n; i < j)$, and $b_{ij} = \bar{b}_{ji}$ $(i > j)$. Also let us set

$$h_1(B) = y_0 + y_1B + \cdots + y_kB^k$$

and

$$h_2(B) = z_0 + z_1B + \cdots + z_lB^l.$$

Thus $h_1(B)$, $h_2(B)$ and $h_1(B)h_2(B)$ are Hermitian matrices over $\mathfrak{R}'$, and thus

$$F = | h_1(B)h_2(B) | - | h_1(B) | \cdot | h_2(B) |$$

is a polynomial in the various indeterminates, with integer coefficients. We shall show that this is actually the zero polynomial. To this end, let us formally replace the x_{ii}, x_{ij}, y_i, z_i by arbitrary real numbers and α, β by arbitrary negative real numbers. By this specialization, $h_1(B)$, $h_2(B)$, and $h_1(B)h_2(B)$

become Hermitian matrices over the algebra of real quaternions[9] and hence F vanishes since our theorem is true for this case. Since F vanishes for every such specialization, it follows readily that $F = 0$ as an element of $\mathfrak{R}'$. Since F has integer coefficients, the relation $F = 0$ remains true if the indeterminates are replaced by arbitrary elements of a commutative ring $\mathfrak{C}$ with unit element, it being understood that an integer m is to be replaced by $m \cdot 1$, where 1 is the unit element of $\mathfrak{C}$. It is clear, however, that by proper replacement of the indeterminates in B by elements of $\mathfrak{C}$, B goes over into A, while $h_1(B)$ goes over into $g_1(A)$ by also replacing y_i by a_i, and $h_2(B)$ becomes $g_2(A)$ upon replacing z_i by b_i. Hence the theorem is established.

Now let A be an element of $\mathcal{K}$ with characteristic polynomial $f(\lambda) = | \lambda - a | = \lambda^n + \cdots + a_n$. Let $g(\lambda) = b_0\lambda^m + b_1\lambda^{m-1} + \cdots + b_m$ be an element of $\mathfrak{C}[\lambda]$, and let us define the *resultant* $\mathcal{R}(f, g)$ of $f(\lambda)$ and $g(\lambda)$ in a formal way by the Sylvester determinant

$$\mathcal{R}(f,g) = \begin{vmatrix} 1 & a_1 & \cdot & \cdot & \cdot & a_n & & & \\ & 1 & a_1 & \cdot & \cdot & \cdot & a_n & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & 1 & a_1 & \cdot & \cdot & \cdot & a_n \\ b_0 & b_1 & \cdot & \cdot & \cdot & b_m & & & \\ & b_0 & b_1 & \cdot & \cdot & \cdot & b_m & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & b_0 & b_1 & \cdot & \cdot & \cdot & b_m \end{vmatrix},$$

the blank spaces consisting of zeros. We may now prove

THEOREM 10. *If $A \in \mathcal{K}$ and $g(\lambda) \in \mathfrak{C}[\lambda]$, then*

$$| g(A) | = \mathcal{R}(f, g). \tag{32}$$

Thus, in particular, if A_1 and A_2 are elements of $\mathcal{K}$ with the same characteristic polynomial, then

$$| g(A_1) | = | g(A_2) |$$

for every element $g(\lambda)$ of $\mathfrak{C}[\lambda]$.

The proof for the commutative case has already been given in [3] and is naturally based on the similar theorem due to Frobenius for the case in which the coefficients are from an algebraically closed field. The proof for quaternion rings can be carried through in the same general way as the proof of Theorem 9 except for the necessity of proving relation (32) for the special case of quaternions over the reals.

Thus, suppose A is a Hermitian matrix over real quaternions and let $g(\lambda)$ have real coefficients. Now $g(\lambda)$ can be factored in the real field into a product of linear or quadratic factors. Since

$$\mathcal{R}(f, g_1g_2) = \mathcal{R}(f, g_1)\mathcal{R}(f, g_2),$$

[9] To get the usual basis we only need to replace w_1 by $w_1/(-\alpha)^{\frac{1}{2}}$, w_2 by $w_2/(-\beta)^{\frac{1}{2}}$ and w_3 by $w_3/(\alpha\beta)^{\frac{1}{2}}$.

it is sufficient, in view of Theorem 9, to prove formula (32) for linear and quadratic $g(\lambda)$. A direct calculation shows it to be true for linear $g(\lambda)$ and therefore also for quadratic $g(\lambda)$ which can be factored into real linear factors. Now set

$$g(\lambda) = z_0 + z_1\lambda + z_2\lambda^2,$$

with indeterminate z_0, z_1, z_2. Then $|\,g(A)\,|$ and $\mathcal{R}(f, g)$ are polynomials in z_0, z_1, z_2 which are equal for all choices of real z_0, z_1, z_2 such that $z_1^2 - 4z_0z_2 \geqq 0$. This implies, however, that $|\,g(A)\,|$ and $\mathcal{R}(f, g)$ are identically equal as polynomials in z_0, z_1, z_2 and thus are equal for all real quadratic $g(\lambda)$. Relation (32) is established for the case in which A is a Hermitian matrix over real quaternions and $g(\lambda)$ has real coefficients. The rest of the proof of Theorem 10 follows by arguments similar to those used in the proof of the preceding theorem.

The following result is fundamental for certain purposes:

THEOREM 11. *If $A \in \mathcal{K}$, the following statements are equivalent:*
(i) $|\,A\,|$ *is a divisor of zero in* $\mathfrak{R}$.
(ii) A *is a divisor of zero in* $\mathfrak{R}_n$.
(iii) A *is a divisor of zero in* $\mathfrak{C}[A]$.

It is obvious that (iii) implies (ii) and furthermore (ii) implies (i) by Theorem 3. The essential part of the theorem is contained in the proof that (i) implies (iii). This has been established in [3] for Case 2 and inasmuch as this proof will apply equally to the first case, we shall not reproduce the proof here. We may remark that if $|\,A\,|$ is a divisor of zero in $\mathfrak{R}$, it is also a divisor of zero in $\mathfrak{C}$. This fact is important in the proof.

An element $g(\lambda)$ of $\mathfrak{C}[\lambda]$ is said to be *prime* to an ideal $\mathfrak{m}$ if $g(\lambda)h(\lambda) \equiv 0\ (\mathfrak{m})$ implies that $h(\lambda) \equiv 0\ (\mathfrak{m})$, it being understood that $h(\lambda) \in \mathfrak{C}[\lambda]$. If $\mathfrak{m}$ is the minimum ideal of the matrix A, it follows from the preceding theorem, together with the fact that $\mathfrak{C}[A] \cong \mathfrak{C}[\lambda]/\mathfrak{m}$, that $g(\lambda)$ is prime to $\mathfrak{m}$ if and only if $|\,g(A)\,|$ is not a divisor of zero in $\mathfrak{C}$. But $|\,g(A)\,| = \mathcal{R}(f, g)$ and it has been shown elsewhere (see [3], p. 170) that $g(\lambda)$ is prime to $(f(\lambda))$ if and only if $\mathcal{R}(f, g)$ is not a divisor of zero in $\mathfrak{C}$. We have therefore proved

THEOREM 12. *If $A \in \mathcal{K}$ has characteristic polynomial $f(\lambda)$ and minimum ideal $\mathfrak{m}$, an element $g(\lambda)$ of $\mathfrak{C}[\lambda]$ is prime to $\mathfrak{m}$ if and only if it is prime to $(f(\lambda))$.*

Now Theorems 5 and 11 show at once that if $A \in \mathcal{K}$, then A has an inverse in $\mathfrak{R}_n$ (actually in $\mathfrak{C}[A]$) if and only if $|\,A\,|$ has an inverse in $\mathfrak{R}$. Since, by the corollary to Theorem 6, if $h(\lambda) \equiv 0\ (\mathfrak{m})$, then $[h(\lambda)]^n \equiv 0\ (f(\lambda))$, it is easy to see that $(h(\lambda), \mathfrak{m}) = (1)$ if and only if $(h(\lambda), f(\lambda)) = (1)$. The following result is then almost obvious.

THEOREM 13. *If $A \in \mathcal{K}$ and $h(\lambda) \in \mathfrak{C}[\lambda]$, then $h(A)$ has an inverse in $\mathfrak{C}[A]$ if and only if in $\mathfrak{C}[\lambda]$ we have $(h(\lambda), f(\lambda)) = (1)$.*

We conclude with an application of this result. If A and B are elements of $\mathcal{K}$ and $AB = BA$, then an examination of the Sylvester identities (see [9],

p. 27) shows that they can be carried over to the present situation. Hence, in particular, there is an expression of the form

$$f'(A)B = g(A),$$

where $f'(\lambda)$ is the formal derivative of $f(\lambda)$ with respect to λ, and $g(\lambda) \epsilon \mathfrak{C}[\lambda]$. If now $(f'(\lambda), f(\lambda)) = (1)$, $f'(A)$ has, by Theorem 13, an inverse in $\mathfrak{C}[A]$. We have therefore proved

THEOREM 14. *If $A \epsilon \mathcal{K}$ and in $\mathfrak{C}[\lambda]$, $(f'(\lambda), f(\lambda)) = (1)$, the only elements of $\mathcal{K}$ commutative with A are the elements of $\mathfrak{C}[A]$.*

BIBLIOGRAPHY

1. A. A. ALBERT, *Modern Higher Algebra*, Chicago, 1937.
2. N. JACOBSON, *An application of E. H. Moore's determinant of a Hermitian matrix*, Bulletin of the American Mathematical Society, vol. 45(1939), pp. 745–748.
3. N. H. MCCOY, *Divisors of zero in matric rings*, ibid., vol. 47(1941), pp. 166–172.
4. N. H. MCCOY, *A generalization of Ostrowski's theorem on matric identities*, ibid., vol. 46 (1940), pp. 490–495.
5. N. H. MCCOY, *Concerning matrices with elements in a commutative ring*, ibid., vol. 45(1939), pp. 280–284.
6. E. H. MOORE AND R. W. BARNARD, *General Analysis*, Part I, Philadelphia, 1935.
7. A. OSTROWSKI, *On a theorem concerning identical relations between matrices*, Quarterly Journal of Mathematics, vol. 9(1938), pp. 241–245.
8. H. B. PHILLIPS, *Functions of matrices*, American Journal of Mathematics, vol. 41(1919), pp. 266–278.
9. J. H. M. WEDDERBURN, *Lectures on Matrices*, American Mathematical Society Colloquium Publications, vol. 17(1934).

SMITH COLLEGE.

CENTRAL CHAINS OF IDEALS IN AN ASSOCIATIVE RING

By S. A. Jennings

In this paper we have two main ideas in mind. It is well known that there is a considerable similarity between those properties of a group associated with its commutator structure and the theory of Lie rings.[1] We show first that, with suitable definitions of "commutator ideal", many of the properties of commutator subgroups have analogues in the theory of associative rings. In particular, we are interested in extending the notions of "nilpotent group" and "solvable group" to rings. On the other hand, every associative ring determines a Lie ring, which we have called the "associated Lie ring" (cf. §6), and the question arises as to how far the solvability or nilpotency of this Lie ring determines corresponding properties of the original associative ring as we have defined them in the first part. In the case of algebras we answer this question completely and show that if the Lie algebra is solvable (or nilpotent) the associative algebra has the corresponding property. For a general ring, however, we obtain only a partial answer.

In seeking an analogue for the commutator subgroup of two given normal subgroups, a difficulty arises at once. As is well known, the subgroup generated by all commutators of a group is a normal subgroup; in a ring the subring generated by all elements of the form $xy - yx$ is not in general an ideal. We overcome this difficulty by defining the "commutator ideal" of two given ideals A, B of the associative ring R as the smallest ideal of R containing all elements of the form $ab - ba$, where $a \in A$, $b \in B$. However, there are some disadvantages to this definition, as compared with the corresponding one in the theory of groups, and in consequence, commutator subgroups enjoy some properties which have no analogues for commutator ideals. In a group, a normal subgroup, besides having the property that its residue classes again form a group, is such that it is transformed into itself by inner automorphisms. Ideals in a ring play no such dual rôle, and in general those properties of commutator subgroups depending on the second of these facts have no analogue in our theory.

1. **Commutators.** Let A and B be any two (not necessarily distinct) ideals of an associative ring R. We define the *commutator ideal*, $A \circ B$, of A and B to be the ideal of R generated by all elements of the form $ab - ba$, where $a \in A$, $b \in B$, that is, $A \circ B$ is the smallest ideal of R containing all elements of the form $ab - ba$.

In what follows, it will be convenient to write $xy - yx$ as $x \circ y$; to avoid the possibility of confusion between this notation and that defined above for com-

Received November 8, 1941. Presented to the American Mathematical Society, September 2, 1941.

[1] For properties of commutator subgroups see [1] and [7]. (Numbers in brackets refer to the bibliography.) The relations between groups and Lie rings are discussed in [5] and [6].

mutator ideals, we shall use lower case letters for individual ring elements and denote ideals by capital letters.

We consider some elementary properties of commutator ideals. Using our definition, any element of $A \circ B$ may be written as a sum of one or more elements of the following types:

$$a \circ b, \qquad x(a \circ b), \qquad (a \circ b)y, \qquad x(a \circ b)y,$$

where a, b belong to A, B respectively, and x, y are in R. For any three elements of R, however, we have the Jacobi identity

$$(x \circ y) \circ z + (y \circ z) \circ x + (z \circ x) \circ y = 0$$

and hence

$$\begin{aligned}(a \circ b)y &= y(a \circ b) + (a \circ b) \circ y \\ &= y(a \circ b) + a \circ (b \circ y) + (a \circ y) \circ b;\end{aligned}$$

since A, B are ideals, $a \circ y \in A$, $b \circ y \in B$, and hence $(a \circ b)y$ may be expressed as a sum of elements of the types $(a \circ b)$, $x(a \circ b)$. Similarly $x(a \circ b)y$ may be written as a sum of elements having the form $x(a \circ b)$. We have proved, therefore, that any element of $A \circ B$ may be written as a sum of elements of one or both of the types

(1A) $$a \circ b; \qquad x(a \circ b).$$

If C is the subring of R generated by all elements $a \circ b$, where $a \in A$, $b \in B$, the remark embodied in (1A) is equivalent to the equality

(1B) $$A \circ B = C \cup_{\circ} (R \cdot C),$$

and if R has a unit element we have

(1C) $$A \circ B = R \cdot C.$$

We have also for any ring R

(1D) $$A \circ B = B \circ A,$$

(1E) $$A \circ B \subseteq A \cdot B \subseteq A \cup B.$$

It should be noted that our definition of commutator ideal depends not only on the ideals A and B but also on the underlying ring R; for example, if S is a subring of R containing both A and B, then the ideal $A \circ B$ taken relative to the ring R will in general be greater than the ideal $A \circ B$ relative to S. This state of affairs does not arise in the theory of groups or Lie rings. For the most part no ambiguity arises, and unless otherwise stated (cf. §5) all commutator ideals are to be taken relative to the whole ring.

2. **Central chains of ideals.** Let

$$(2A) \qquad R = M_1 \supseteqq M_2 \supseteqq \cdots \supseteqq M_m \supseteqq M_{m+1} = 0$$

be a chain of ideals of R. The chain (2A) will be called a *central chain of ideals*[2] if we have

$$(2B) \qquad R \circ M_i \subseteq M_{i+1} \qquad (i = 1, 2, \cdots, m).$$

If the situation is as in (2A) with M_{m+1} equal to the zero ideal, we say m is the (formal) length of the chain. It will be convenient to define $M_{m+2} = M_{m+3} = \cdots = 0$ for any chain of ideals. The condition (2B) is equivalent to the condition that M_i/M_{i+1} belong to the center of R/M_{i+1} $(i = 1, 2, \cdots)$.

Rings which possess central chains of ideals have special properties; we investigate some of them by considering a particular central chain. For any ring R we may form a descending chain of ideals $R = H_1 \supseteqq H_2 \supseteqq \cdots$ by setting

$$(2C) \qquad R = H_1; \qquad H_{\rho+1} = H_\rho \circ R \text{ for } \rho \geqq 1.$$

We say that the ring R is of *finite class* if $H_m = 0$ for some m; if the situation is as follows

$$(2D) \qquad R = H_1 \supset H_2 \supset \cdots \supset H_c \supset H_{c+1} = 0,$$

then c will be said to be the *class* of R. The chain (2D) will be called the *lower central chain* of R; because of (2C) it is clear that the lower central chain is a central chain as defined above, and hence every ring of finite class has a central chain of ideals.

THEOREM 2.1. *A necessary and sufficient condition that a ring R have a central chain of ideals is that R be of finite class. The length of any central chain of R is at least equal to the class of R, and if*

$$R = M_1 \supseteqq M_2 \supseteqq \cdots \supseteqq M_{m+1} = 0$$

is any central chain, and

$$R = H_1 \supset H_2 \supset \cdots \supset H_c \supset H_{c+1} = 0$$

is the lower central chain of R, we have

$$M_i \supseteqq H_i \qquad (i = 1, 2, \cdots).$$

Proof. Suppose that R has a central chain

$$R = M_1 \supseteqq M_2 \supseteqq \cdots \supseteqq M_{m+1} = 0;$$

then since $H_1 = M_1 = R$, we can suppose that $H_{i-1} \subseteq M_{i-1}$, and show that this implies $H_i \subseteq M_i$. However, since, by (2C) and (2B),

$$H_i = (H_{i-1} \circ R) \subseteq (M_{i-1} \circ R) \subseteq M_i,$$

[2] For the corresponding concept in the theory of groups see [1] or [7].

we have

$$H_i \subseteq M_i$$

as required. Now since $M_{m+1} = 0$, $H_{m+1} = 0$, and R is of finite class. The remainder of the theorem follows at once.

The following theorems show that rings of finite class are fairly common.

THEOREM 2.2. *A necessary and sufficient condition that a ring R be commutative is that R be of class one, that is, $H_2 = 0$.*

Proof. If R is commutative, then $xy - yx = 0$ for all $x, y \in R$, and hence, by (1A) and (2C), $H_2 = 0$. Conversely, if $H_2 = 0$, all expressions of the form $xy - yx$ must vanish, and hence R is commutative.

For any ring R, the ideal $R \circ R (= H_2)$ will be called the *derived ring*. It is characterized in the following

THEOREM 2.3. *The derived ring of R is the intersection of all ideals of R whose factor rings are commutative; $R/(R \circ R)$ is commutative and if A is any ideal of R containing $(R \circ R)$, then R/A is commutative.*

Proof. If R/A is commutative then $xy - yx \equiv 0$ modulo A, and hence $(R \circ R) \subseteq A$. Conversely, if $(R \circ R) \subseteq A$ then $xy - yx \equiv 0$ modulo A, and R/A is commutative.

THEOREM 2.4. *Every nilpotent ring N is of finite class, and the class is at most equal to the exponent of N. In particular, the powers of N*

$$N = N^1 \supset N^2 \supset \cdots \supset N^\lambda \supset N^{\lambda+1} = 0$$

form a central chain of ideals of N with the stronger property

$$N^\rho \circ N^\sigma \subseteq N^{\rho+\sigma}.$$

Proof. If N has exponent λ, then from (1E) and (2C) we have $H_\rho \subseteq N^\rho$ and hence $H_{\lambda+1} = 0$. Similarly $N^\rho \circ N^\sigma \subseteq N^\rho \cdot N^\sigma \subseteq N^{\rho+\sigma}$.

3. **Properties of the lower central chain.** We consider now some further properties of central chains, and in particular, of the lower central chain.[3] Unless otherwise mentioned, H_i will denote the i-th member of the lower chain of an arbitrary ring R of finite class. To stress the fact that H_i belongs to the ring R, we sometimes write $H_i = H_i(R)$.

THEOREM 3.1. *If $R = M_1 \supseteqq M_2 \supseteqq \cdots \supseteqq M_{m+1} = 0$ is any central chain of ideals of R, and $R = H_1 \supset H_2 \supset \cdots \supset H_{c+1} = 0$ is the lower central chain, then for $\rho, \sigma = 1, 2, \cdots$*

$$H_\rho \cdot M_\sigma \subseteq M_{\rho+\sigma-1},$$

$$M_\sigma \cdot H_\rho \subseteq M_{\rho+\sigma-1}.$$

[3] The theorems in this section, with the exception of (3.1), (3.3) and (3.9), have analogues in the theory of groups.

THEOREM 3.2. *Under the same hypothesis as* (3.1),

$$H_\rho \circ M_\sigma \subseteq M_{\rho+\sigma} \qquad (\rho, \sigma = 1, 2, \cdots).$$

We prove first (3.1) and use it to establish (3.2).

Since M_σ is an ideal and $H_1 = R$, we have $H_1 \cdot M_\sigma \subseteq M_\sigma$, and hence (3.1) holds for $\rho = 1$ and all σ. We use induction and suppose that $H_\tau \cdot M_\sigma \subseteq M_{\tau+\sigma-1}$ for $\tau \leqq \rho - 1$ and all σ; we must show that this implies that $H_\rho \cdot M_\sigma \subseteq M_{\rho+\sigma-1}$ for all σ. Let a be any element of R, $h_{\rho-1}$ any element of $H_{\rho-1}$, and m_σ any element of M_σ. Then $h_{\rho-1} \circ a$ is in H_ρ. Consider the element $(h_{\rho-1} \circ a)m_\sigma$. Using the identity

$$(x \circ y)z = x(y \circ z) + (xz) \circ y,$$

we have

$$(h_{\rho-1} \circ a)m_\sigma = h_{\rho-1}(a \circ m_\sigma) + (h_{\rho-1} \cdot m_\sigma) \circ a.$$

Now $a \circ m_\sigma \in M_{\sigma+1}$, and hence by our induction $h_{\rho-1} \cdot (a \circ m_\sigma) \in M_{\rho+\sigma-1}$; again, $h_{\rho-1} \cdot m_\sigma \in M_{\rho+\sigma-2}$ and hence $(h_{\rho-1} \cdot m_\sigma) \circ a \in M_{\rho+\sigma-1}$, which proves that $(h_{\rho-1} \circ a)m_\sigma \in M_{\rho+\sigma-1}$. Now by (1A) and (2C) every element of H_ρ may be written as a sum of elements of the forms $(h_{\rho-1} \circ a)$, $x(h_{\rho-1} \circ a)$, where x runs over all elements of R, and it follows that $H_\rho \cdot M_\sigma \subseteq M_{\rho+\sigma-1}$, as required. Similarly, using the identity

$$z(x \circ y) = (zx) \circ y + (y \circ z)x$$

we may prove that $M_\sigma \cdot H_\rho \subseteq M_{\rho+\sigma-1}$.

To prove (3.2) we note that for all σ, $H_1 \circ M_\sigma \subseteq M_{\sigma+1}$, since M_σ has the property (2B). We suppose, therefore, that $H_\tau \circ M_\sigma \subseteq M_{\tau+\sigma}$ for $\tau \leqq \rho - 1$ and all σ, and proceed as above. Using the same notation as before consider $(h_{\rho-1} \circ a) \circ m_\sigma$. The Jacobi identity gives

$$(h_{\rho-1} \circ a) \circ m_\sigma = h_{\rho-1} \circ (a \circ m_\sigma) + a \circ (m_\sigma \circ h_{\rho-1})$$

and hence by our induction $(h_{\rho-1} \circ a) \circ m_\sigma$ belongs to $M_{\rho+\sigma}$. For brevity, set $c_\rho = (h_{\rho-1} \circ a)$; our theorem will be proved if we show that every element having the form $(xc_\rho) \circ m_\sigma$ belongs to $M_{\rho+\sigma}$. We have identically

$$(xc_\rho) \circ m_\sigma = x(c_\rho \circ m_\sigma) + (x \circ m_\sigma)c_\rho .$$

Since we have shown $(c_\rho \circ m_\sigma) \in M_{\rho+\sigma}$, $x(c_\rho \circ m_\sigma) \in M_{\rho+\sigma}$. Now $(x \circ m_\sigma) \in M_{\sigma+1}$, and hence, using (3.1), $(x \circ m_\sigma)c_\rho \in M_{\rho+\sigma}$, and hence $(xc_\rho) \circ m_\sigma \in M_{\rho+\sigma}$, and we have our theorem.

As immediate corollaries of the above we state

THEOREM 3.3. $H_\rho \cdot H_\sigma \subseteq H_{\rho+\sigma-1}$.

THEOREM 3.4. $H_\rho \circ H_\sigma \subseteq H_{\rho+\sigma}$.

We consider now some further properties of the commutator ideal $A \circ B$. The following lemmas are readily verified.

LEMMA 3.5. *Let the ring R be mapped homomorphically on the ring $\bar{R}$, $R \to \bar{R}$, and let A and B be any two ideals of R and $\bar{A}$, $\bar{B}$ the corresponding ideals in R such that $A \to \bar{A}$, $B \to \bar{B}$. Then in this homomorphism any element of $A \circ B$ is mapped upon an element of $\bar{A} \circ \bar{B}$.*

LEMMA 3.6. *Under the same assumptions as in (3.5), let M be the ideal in R such that $R/M \cong \bar{R}$. Then the ideal $(A \circ B) \cup M$ maps completely into the ideal $\bar{A} \circ \bar{B}$ in R, that is, every element of $\bar{A} \circ \bar{B}$ is the image of an element in $(A \circ B) \cup M$.*

As consequences of (3.5) and (3.6) we have

THEOREM 3.7. *If M is any ideal of R such that R/M is of class $\rho - 1$, then $M \supseteq H_\rho$; in other words, H_ρ is the smallest ideal in R whose factor ring is of class $\rho - 1$.*

THEOREM 3.8. *If M is any ideal in R, then $H_\rho \cup M$ is the ideal in R which maps completely upon the ideal $H_\rho(R/M)$ in the homomorphism of R upon R/M.*

For later use we state the following consequence of (3.3).

THEOREM 3.9. *If R is of finite class the ideal $R \circ R$ is nilpotent.*

Proof. Since $R \circ R = H_2$, we have $H_2 \cdot H_2 \subseteq H_3$, and in general $H_2^k \subseteq H_{k+1}$, and hence $H_2^c = 0$, which proves that H_2 is nilpotent, of exponent at most $c - 1$.

4. **The upper central chain.** For certain types of rings of finite class we may define also the *upper central chain* of ideals. For the rest of this section, let R be a ring of class c with ascending chain condition for ideals.[4] Consider the ideal H_c; since $H_c \circ R = H_{c+1} = 0$, H_c belongs to the center of R, which is therefore not zero. We see also that every ring of finite class contains at least one "central ideal",[5] namely, H_c. Let Z_1 denote the maximal central ideal of R. If we set Z_0 equal to the 0-ideal, we define the *upper central chain* of R

$$0 = Z_0 \subset Z_1 \subset Z_2 \subset \cdots \subset Z_{c'} = R \tag{4A}$$

by setting Z_i equal to the ideal in R which maps upon the ideal $Z_1(R/Z_{i-1})$ of R/Z_{i-1}, that is, Z_i is the maximal ideal of R contained in the center of R modulo Z_{i-1}. It is clear that the chain (4A) is a central chain, since if a is any element of R and z_i belongs to Z_i, then $az_i - z_ia \equiv 0$ modulo Z_{i-1}, and hence $R \circ Z_i \subseteq Z_{i-1}$.

Let

$$R = M_1 \supseteq M_2 \supseteq \cdots \supseteq M_m \supseteq M_{m+1} = 0$$

be any central chain of ideals; then since by (2B) M_m is a central ideal, we have $M_m \subseteq Z_1$. Let us assume that we know that $M_{m-i+2} \subseteq Z_{i-1}$ and show that this implies that $M_{m-i+1} \subseteq Z_i$. We have $M_{m-i+1} \circ R \subseteq M_{m-i+2} \subseteq Z_{i-1}$ and

[4] If the elements of R are well ordered, the assumption of the maximal condition is not necessary for the definition of the upper central chain.

[5] A central ideal is an ideal contained in the center.

hence M_{m-i+1} belongs to the center of R modulo Z_{i-1} and hence is contained in Z_i. We have proved that

$$Z_i \supseteqq M_{m-i+1} \qquad (i = 1, 2, \cdots). \tag{4B}$$

It follows that the length, c', of the upper central chain is less than or equal to the length of any central chain, and hence, by (2.1), must equal the class c of the ring. We have proved, therefore, the following

THEOREM 4.1. *If R is a ring of finite class with ascending chain condition for two sided ideals, the upper central chain*

$$R = Z_c \supset Z_{c-1} \supset \cdots \supset Z_1 \supset Z_0 = 0$$

exists and has the property that, if

$$R = M_1 \supseteqq M_2 \supseteqq \cdots \supseteqq M_m \supseteqq M_{m+1} = 0$$

is any central chain of R, then

$$Z_i \supseteqq M_{m-i+1} \qquad (i = 1, 2, \cdots).$$

In particular, $Z_i \supseteqq H_{c-i+1}$.

It will be convenient to define $Z_i = 0$ for $i = -1, -2, \cdots$. Then using (3.1) and (3.2) we have at once the following two theorems.

THEOREM 4.2. $H_i \cdot Z_j \subseteq Z_{j-i+1}$; $Z_j \cdot H_i \subseteq Z_{j-i+1}$. *In particular,* $H_i \cdot Z_{i-1} = Z_{i-1} \cdot H_i = 0$.

THEOREM 4.3. $H_i \circ Z_j \subseteq Z_{j-i}$. *In particular,* $H_i \circ Z_i = 0$.

The following result illustrates the close analogy between the properties of the upper central chain and those of the upper central series of a group (cf. [7], p. 108, Theorem 14).

THEOREM 4.4. *If the ideal N is contained in Z_{i+1}, but not contained in Z_i, then the chain*

$$N \supset N \cap Z_i \supset N \cap Z_{i-1} \supset \cdots \supset N \cap Z_1 \supset 0$$

is a central chain of N, and is strictly decreasing.

Proof. The chain is obviously a central chain: we must show that it is strictly decreasing. We have

$$R \circ N \subseteq N \cap (R \circ Z_{i+1}) \subseteq N \cap Z_i .$$

Since N is not contained in Z_i, $R \circ N$ is not contained in Z_{i-1} and hence $N \cap Z_i$ is not contained in $N \cap Z_{i-1}$; repeating this argument using $N \cap Z_i$, instead of N, etc., we see that the chain is strictly decreasing.

The following is an immediate consequence of (4.4).

THEOREM 4.5. *Any ideal of a ring of finite class contains elements of the maximal central ideal Z_1.*

It is convenient at this point to mention the notion of *centralisor ideal*. Let A be any ideal of R; then the centralisor ideal of A, $C(A)$, may be defined as the maximal ideal of R such that $C(A) \circ A = 0$. If B is any ideal such that $A \circ B = 0$, then $B \subseteq C(A)$, and alternatively we may define $C(A)$ as the union of all such ideals. The maximal central ideal Z_1 could be defined as the centralisor of R, and similarly for the other terms of the upper central chain.

5. **Solvable rings.** We have defined the derived ring $R \circ R$ of the given ring R, and have seen that it is the intersection of all ideals of R whose factor rings are commutative. It is convenient to denote the derived ring by $R^{(1)}$, and we set $R = R^{(0)}$. We may form the derived ring of $R^{(1)}$, which we denote by $R^{(2)}$ and in general form $R^{(n)}$, the derived ring of $R^{(n-1)}$. We obtain in this way a chain of *subrings*

$$R = R^{(0)} \supseteqq R^{(1)} \supseteqq R^{(2)} \supseteqq \cdots \supseteqq R^{(p)} \supseteqq \cdots$$

such that $R^{(n)}$ is the derived ring of $R^{(n-1)}$ $(n = 1, 2, \cdots)$. If for some integer k we have $R^{(k)} = 0$ the ring R is said to be *solvable* and we call the chain

$$R = R^{(0)} \supset R^{(1)} \supset R^{(2)} \supset \cdots \supset R^{(k)} = 0 \tag{5A}$$

the *derived chain* of R. We note that while $R^{(n)}$ is an ideal of $R^{(n-1)}$, it may not be an ideal of R.

We prove first the following

THEOREM 5.1. *A necessary and sufficient condition that the ring R be solvable is that there exist a chain of subrings of R*

$$R = A_0 \supseteqq A_1 \supseteqq A_2 \supseteqq \cdots \supseteqq A_{m-1} \supseteqq A_m = 0$$

terminating with zero, such that

(1) *A_p is an ideal of A_{p-1} $(p = 1, 2, \cdots)$,*

(2) *A_{p-1}/A_p is commutative $(p = 1, 2, \cdots)$.*

Proof. The necessity follows since the derived chain of a solvable ring satisfies (1) and (2). Conversely, the conditions are sufficient, because, using (2.3), we see that $A_p \supseteqq R^{(p)}$ $(p = 1, 2, \cdots)$ and hence $R^{(m)} = 0$, if $A_m = 0$.

It follows at once that if R is a ring with a central chain of ideals then these ideals satisfy (5.1), (1) and (5.1), (2), and we have, therefore,

THEOREM 5.2. *Every ring of finite class is solvable.*

We prove now a result due to Jacobson [3] which we shall need later.

LEMMA 5.3. *Let R be any associative ring and N a nilpotent subring of R which is closed under commutation with elements of R, that is, if $x \in R$, $n \in N$, then $x \circ n \in N$. Then $R \cdot N \cup N$ is a nilpotent ideal of R.*

Proof. It is clear that $\bar{N} = RN \cup N$ is an ideal since $NR \subseteq RN \cup N$. Consider

$$\begin{aligned}\bar{N}^2 &= (RN \cup N)(RN \cup N)\\ &\subseteq (RNRN) \cup (NRN) \cup (RN^2) \cup N^2.\end{aligned}$$

Now since N is closed under commutation,

$$NRN \subseteq RN^2 \cup N^2$$

and hence

$$\bar{N}^2 \subseteq RN^2 \cup N^2.$$

Similarly

$$\bar{N}^k \subseteq RN^k \cup N^k$$

and hence, since $N^k = 0$ for some k, $\bar{N}^k = 0$, which proves that $\bar{N}$ is nilpotent, as required.

We need also the following

LEMMA 5.4. *Let C be a commutative ideal of R. Then $(R \circ C)\cdot C = 0$, and in particular, $(R \circ C)\cdot(R \circ C) = 0$, that is, $(R \circ C)$ is nilpotent.*

Proof. If C is commutative, then $C \circ C = 0$, and hence, if c, c' are any two elements of C, and $x \in R$, we have $xc \in C$ and therefore

$$(xc) \circ c' = 0.$$

However, $(xc) \circ c' = x(c \circ c') + (x \circ c')c$, and hence $(x \circ c')c = 0$, which shows that $(R \circ C)\cdot C = 0$. Since $C \supseteqq R \circ C$, we have $(R \circ C)\cdot(R \circ C) = 0$, that is, $R \circ C$ is nilpotent.

We are now in a position to prove the principal result of this section, which is

THEOREM 5.5. *If R is a solvable ring, its derived ring is nilpotent.*

Proof. Suppose first that R is such that $R^{(2)} = 0$. Then $R^{(1)}$ is a commutative ideal of R. We form $R \circ R^{(1)} = S$, say. If $S = 0$, then R is a ring of class 2, and hence $R^{(1)}(= H_2(R))$ is nilpotent by (3.9). If $S \neq 0$, then by (5.4) S is a nilpotent ideal of R. Moreover, R/S is of class 2, and hence $R^{(1)}/S(= H_2(R/S))$ is nilpotent, and since both $R^{(1)}/S$ and S are nilpotent, so is $R^{(1)}$, which proves our theorem for the case that $R^{(2)} = 0$.

Suppose now that we have

$$R = R^{(0)} \supset R^{(1)} \supset \cdots \supset R^{(k)} = 0$$

with $k > 2$. We proceed by induction and suppose (5.5) true for rings with derived chain of length less than k. In particular, (5.5) is true for the ring $R^{(1)}$ whose derived chain is of length $k - 1$, and hence $R^{(2)}$ is a nilpotent ideal of $R^{(1)}$, and hence is a nilpotent subring of R. We show first that $R^{(2)}$ is closed under commutation with elements of R; $R^{(1)}$ is closed under commutation since it is an ideal of R. Any element of $R^{(2)}$ may be written as a sum of elements of one or both of the types

$$a_1 \circ b_1, \qquad c_1(a_1 \circ b_1),$$

where a_1, b_1, c_1 are elements of $R^{(1)}$. Consider $(a_1 \circ b_1) \circ x = a_1 \circ (b_1 \circ x) + b_1 \circ (x \circ a_1)$; since $R^{(1)}$ is an ideal, $b_1 \circ x$ and $x \circ a_1$ are again in $R^{(1)}$, and hence $(a_1 \circ b_1) \circ x$ is in $R^{(2)}$. Similarly, since

$$[c_1(a_1 \circ b_1)] \circ x = c_1[(a_1 \circ b_1) \circ x] + (c_1 \circ x)(a_1 \circ b_1)$$

and $(a_1 \circ b_1) \circ x \in R^{(2)}$ and $(c_1 \circ x) \in R^{(1)}$, $[c_1(a_1 \circ b_1)] \circ x$ is in $R^{(2)}$ and hence $R^{(2)}$ is closed under commutation.

We form the ideal

$$M = R^{(2)} \cup RR^{(2)}$$

of R. Since $R^{(2)}$ is nilpotent and closed under commutation, by (5.3) M is a nilpotent ideal of R. Consider R/M; its second derived ring is zero, since $R^{(2)} \subseteq M$, and hence $R^{(1)}/M$ is nilpotent by our first case. It follows that $R^{(1)}$ is nilpotent since $R^{(1)}/M$ and M are, which proves our theorem.

An equivalent statement to the above is

THEOREM 5.6. *A necessary and sufficient condition that a ring R be solvable is that R contain a nilpotent ideal N such that R/N is commutative. In particular, if R has a radical, it is necessary and sufficient that R be commutative modulo its radical.*

Proof. The necessity of the condition follows; to see that it is sufficient we use (5.1). Suppose that N is a nilpotent ideal of R. We form the powers of N, which are ideals of R, and consider the chain

$$R \supset N^1 \supset N^2 \supset \cdots \supset N^\lambda \supset 0. \tag{5B}$$

Now R/N is commutative, and since N^i/N^{i+1} is commutative, $i = 1, 2, \cdots$, by (5.1) R is solvable.

For some purposes it is inconvenient that the members of the derived chain of a solvable ring are not in general ideals of the ring. We see that it is possible to overcome this difficulty. For a given ring we form a chain of ideals:

$$R = R_0 \supseteqq R_1 \supseteqq R_2 \supseteqq \cdots \supseteqq R_n \supseteqq \cdots, \tag{5C}$$

where $R_1 = R \circ R$, and in general $R_n = R_{n-1} \circ R_{n-1}$, the commutator ideal $R_{n-1} \circ R_{n-1}$ being formed in this case with respect to the whole ring R, and hence R_n being an ideal of R. Alternatively we may define R_n as the intersection of all ideals S of R contained in R_{n-1} such that R_{n-1}/S is commutative. Now if R is solvable, (5B) is a chain of ideals of R such that $N^i \supseteqq R_i$ and hence $R_{\lambda+1} = 0$, and therefore the chain (5C) terminates with the zero ideal. Conversely, if (5C) terminates with zero by (5.1) R is solvable.

We embody these remarks in

THEOREM 5.7. *A ring R is solvable if and only if there exist chains of ideals of R*

$$R = S_0 \supseteqq S_1 \supseteqq S_2 \supseteqq \cdots \supseteqq S_m = 0$$

such that S_{p-1}/S_p is commutative, $p = 1, 2, \cdots, m$.

The following condition for a ring to be solvable is obtained from (5.3).

THEOREM 5.8. *Let C be the subring of R generated by all elements of the form $x \circ y$, where $x, y \in R$. A necessary and sufficient condition that R be solvable is that C be nilpotent.*

Proof. The condition is necessary since $C \subseteq R^{(1)}$, and if R is solvable, $R^{(1)}$ is nilpotent. Conversely, suppose C is nilpotent; from its definition C is closed under commutation with elements of R, and hence by (5.3), the ideal

$$M = C \cup RC$$

is nilpotent. Moreover, R/M is clearly commutative, and hence by (5.6) R is solvable.

6. **The associated Lie ring.** We call the Lie ring formed from a given associative ring R by combining the elements of R under addition and commutation, where the commutator $x \circ y$ of two elements $x, y \in R$ is defined as $xy - yx$, the *Lie ring associated with R.*[6] We shall denote this Lie ring by $\mathfrak{R}$; we may remark that there is no confusion in supposing that R and $\mathfrak{R}$ have the same elements, and we shall use $x \circ y$ to denote the expression $xy - yx$ in R as before. We are concerned in this section with relations between the properties of R and $\mathfrak{R}$.

We recall that for any two ideals $\mathfrak{A}$, $\mathfrak{B}$ of an arbitrary Lie ring $\mathfrak{L}$ we may form the product $[\mathfrak{A}, \mathfrak{B}]$ consisting of all elements of the form $a \circ b$, where $a \in \mathfrak{A}$, $b \in \mathfrak{B}$. The product $[\mathfrak{A}, \mathfrak{B}]$ is again an ideal of $\mathfrak{L}$. A Lie ring $\mathfrak{L}$ is said to be *nilpotent* if the chain

$$\mathfrak{L} = \mathfrak{L}_1 \supset \mathfrak{L}_2 \supset \mathfrak{L}_3 \supset \cdots \supset \mathfrak{L}_\gamma \supset \mathfrak{L}_{\gamma+1} = 0$$

terminates with zero, where $\mathfrak{L}_\rho = [\mathfrak{L}_{\rho-1}, \mathfrak{L}]$. A Lie ring $\mathfrak{L}$ is solvable if the chain

$$\mathfrak{L} = \mathfrak{L}^{(0)} \supset \mathfrak{L}^{(1)} \supset \mathfrak{L}^{(2)} \supset \cdots \supset \mathfrak{L}^{(k)} = 0,$$

where $\mathfrak{L}^{(i)} = [\mathfrak{L}^{(i-1)}, \mathfrak{L}^{(i-1)}]$ terminates with zero.

Suppose now that R is a solvable ring, with derived chain

$$R = R^{(0)} \supset R^{(1)} \supset \cdots \supset R^{(k)} = 0.$$

Then if a_i is any element of $\mathfrak{R}^{(i)}$, it is clear that $a_i \in R^{(i)}$, and hence $\mathfrak{R}^{(k)} = 0$, that is, $\mathfrak{R}$ is a solvable Lie ring. Similarly if R is of finite class and has the lower central chain

$$R = H_1 \supset H_2 \supset \cdots \supset H_c \supset H_{c+1} = 0,$$

then if b_ρ is any element of $\mathfrak{R}_\rho$, $b_\rho \in H_\rho$, and hence $\mathfrak{R}_{c+1} = 0$, that is, $\mathfrak{R}$ is nilpotent, and its class is at most equal to the class of R.

We may state, therefore,

THEOREM 6.1. *If R is a solvable ring with $R^{(k)} = 0$, then the associated Lie ring $\mathfrak{R}$ is solvable, and $\mathfrak{R}^{(k)} = 0$.*

THEOREM 6.2. *If R is of finite class c, then the associated Lie ring $\mathfrak{R}$ is nilpotent and has class at most c.*

[6] For a brief discussion of some of the properties of Lie algebras and Lie rings, and of the Lie ring associated with a given associative ring, cf. [2], [3], [4].

The converse problem to the above, namely, knowing tht $\mathfrak{R}$ is solvable, or nilpotent, to draw conclusions regarding R, is more difficult. For a general ring, our only result along this line is embodied in (6.5), but for the special case when R is an algebra, we prove that, provided the characteristic of the underlying field is not two, the solvability of the associated Lie algebra implies the corresponding property of the original associative algebra, while if the associated Lie algebra is nilpotent, the original algebra is of finite class, whatever the characteristic.

We proceed to prove, first,[7]

THEOREM 6.3. *Let R be an algebra over a field Φ. If the characteristic of Φ is not two, then R is solvable if $\mathfrak{R}$ is solvable. There are non-solvable algebras over any field of characteristic two whose associated Lie algebras are solvable.*

Proof. Let N be the radical of R, and set $S = R/N$. Because of (5.6), it will suffice to show that S is commutative. In what follows, if A is any algebra we denote its associated Lie algebra by A_l. Now if R_l is a solvable Lie algebra, so is S_l. Moreover, S is semi-simple and if

$$S = A_1 \oplus A_2 \oplus A_3 \oplus \cdots \oplus A_t$$

is the splitting of S into simple components, then

$$S_l = (A_1)_l \oplus (A_2)_l \oplus (A_3)_l \oplus \cdots \oplus (A_t)_l$$

and $(A_i)_l$ is solvable, $i = 1, 2, \cdots, t$. We may therefore reduce our problem to the case of a simple algebra A, such that A_l is a solvable Lie algebra, and we wish to prove that A is commutative. Now let Σ be the center of A, and consider the algebra (A over Σ). Clearly A_l is solvable if and only if $(A \text{ over } \Sigma)_l$ is, and hence we may further reduce the problem to the case where A is a normal simple algebra over the field Σ. Let Ω be the algebraic closure of Σ and consider (A over Ω); we know that

$$(A \text{ over } \Omega) \cong \Omega_n,$$

where Ω_n is a complete matric algebra of degree n over Σ, where n is the rank of A over Σ. Now $(A \text{ over } \Omega)_l$ is solvable, since $(A \text{ over } \Sigma)_l$ is and hence $(\Omega_n)_l$ must be solvable. The structure of $(\Omega_n)_l$ is known, viz. [2], pp. 216–217. If the characteristic of Ω (and hence of Φ) is not two, $(\Omega_n)_l$ is solvable if and only if $n = 1$, and if the characteristic is two, if and only if $n = 1$ or 2. We see, therefore, that a simple algebra has a solvable associated Lie algebra if and only if it is of rank one over its center, that is, if it is commutative, provided the characteristic of the field Φ is not two, which proves the first part of our theorem.

Direct calculation shows that the complete matric algebra Φ_2 of degree 2 over any field Φ of characteristic 2 is not solvable, and indeed is such that $(\Phi_2) \circ (\Phi_2) =$

[7] I am indebted to the referee for the simple proof of Theorem 6.3 which follows, and particularly for pointing out to me the exceptional case which arises when the characteristic of the underlying field is 2.

Φ_2. However, $(\Phi_2)_l$ is solvable. It is readily verified that $(\Phi_2)_l$, while solvable, is not a nilpotent Lie algebra and hence in the above, if the condition that the Lie algebra $(A)_l$ be solvable is replaced by the stronger condition that it be nilpotent, then we may conclude that $n = 1$ even in the case of characteristic two, and hence A coincides with its center.

We have proved, therefore,

THEOREM 6.4. *Any semi-simple algebra whose associated Lie algebra is nilpotent is commutative.*

In general, little seems to be known of the conditions which must be imposed on an associative ring R in addition to the condition that $\mathfrak{R}$ be solvable (or nilpotent) in order to ensure that R enjoy the corresponding property. We prove here that if R is already known to be solvable, then the nilpotency of $\mathfrak{R}$ implies that R is of finite class.

THEOREM 6.5. *A solvable ring R is of finite class if and only if its associated Lie ring $\mathfrak{R}$ is nilpotent.*

Proof. We have seen that if R has finite class, so has $\mathfrak{R}$. Conversely, suppose that $\mathfrak{R}$ has class γ; if $\gamma = 1$, R is commutative, and hence has class 1, and our theorem is true. Suppose, therefore, that our theorem holds if $\mathfrak{R}$ has class at most $\gamma - 1$. Let C be the subring of R generated by those elements of R which belong to $\mathfrak{R}_\gamma$; then since all such elements belong to the center of R, $c \circ x = 0$, where $c \in C$ and $x \in R$. Hence C is closed under commutation, and since $C \in R^{(1)}$, and $R^{(1)}$ is nilpotent by (5.6), C is nilpotent, and hence, by (5.3)

$$M = C \cup RC$$

is a nilpotent ideal of R. Consider R/M; it is a solvable ring, and the class of its associated Lie ring is at most $\gamma - 1$; hence by our induction R/M is of finite class, and therefore for some integer ρ, $H_\rho(R) \subseteq M$. Our theorem will be proved if we show that, for some integer σ,

$$P_\sigma \equiv (\cdots((M \circ R) \circ R) \circ \cdots) = 0,$$

where the above expression is obtained by commuting M with R σ times, for then $H_{\rho+\sigma}(R) \subseteq P_\sigma = 0$.

We consider, therefore, $M \circ R$. Let m' be any element of M and $x \in R$; if $m' \in C$, then $m' \circ x = 0$. If $m' \in RC$, m' may be written as a sum of elements of the form $m = yc$, where $y \in R$, $c \in C$, and we have

$$m \circ x = (yc \circ x) = (y \circ x)c$$

since c is in the center of R. It follows that $M \circ R \subseteq H_2 \cdot C$, where $H_2 = H_2(R)$. If $M \circ R \neq 0$, consider $(M \circ R) \circ R$; we see similarly that $(M \circ R) \circ R \subseteq H_3 \cdot C$ and in general

$$P_\tau \subseteq H_{\tau+1} \cdot C.$$

Now we know that $H_\rho \subseteq M$, and hence

$$P_{\rho-1} \subseteq M \cdot C \subseteq M^2 \subseteq RC^2 \cup C^2.$$

Similarly $P_\rho \subseteq H_2 \cdot C^2$, and if we continue in this way we see, since C is nilpotent, that there exists an integer σ such that $P_\sigma = 0$, and our theorem is proved.

We are now in a position to prove that every algebra whose associated Lie ring is nilpotent is of finite class. Let R be an algebra such that $\mathfrak{R}$ is nilpotent. Then by (6.4) R/N is commutative, where N is the radical of R, and hence R is solvable, by (5.6). Hence we see, from (6.5), that R is of finite class. We have proved, therefore,

THEOREM 6.6. *Let R be an algebra, and $\mathfrak{R}$ its associated Lie algebra. Then R is of finite class if and only if $\mathfrak{R}$ is nilpotent.*

In the case of algebras, therefore, our concept of finite class is equivalent to the nilpotency of the associated Lie algebra.

7. Solvable nil-rings.

A ring R is said to be a nil-ring if for any $x \in R$ there exists an integer ρ (which may depend on x) such that $x^\rho = 0$. Any nilpotent ring is also a nil-ring, but examples are known of nil-rings which are not nilpotent.

We need the following lemma.

LEMMA 7.1. *A commutative nil-ring generated by a finite number of elements is nilpotent.*

Proof. If R is commutative, and is generated by a finite number of elements, say $a_1, a_2, \cdots, a_r$, then every element of R may be written as a sum of elements of the form

$$a_1^{\alpha_1} \cdot a_2^{\alpha_2} \cdot \cdots \cdot a_r^{\alpha_r} \qquad (\alpha_i \geqq 0; i = 1, 2, \cdots, r).$$

If R is a nil-ring, there exist positive integers $\beta_1, \beta_2, \cdots, \beta_r$ so that

$$a_i^{\beta_i} = 0 \qquad (i = 1, 2, \cdots, r).$$

It follows at once that if $\sigma = \sum \beta_i$, then $R^\sigma = 0$, and R is nilpotent.

In connection with (7.1) we may remark that some condition similar to that in the lemma is essential, since it is easy to construct a nil-ring with an infinite generating set, which is commutative but not nilpotent.

The following theorem may now be proved.

THEOREM 7.2. *Let R be a nil-ring generated by a finite number of elements. A necessary and sufficient condition that R be nilpotent is that R be solvable.*

Proof. If R is solvable, $R^{(1)}$ is a nilpotent ideal of R and $R/R^{(1)}$ is commutative, by (5.5). The factor ring $R/R^{(1)}$, however, is a nil-ring generated by a finite number of elements (since R is) and hence by (7.1) is nilpotent. Since $R/R^{(1)}$ and $R^{(1)}$ are nilpotent, so is R.

Bibliography

1. P. Hall, *A contribution to the theory of groups of prime-power orders*, Proceedings of the London Mathematical Society, (2), vol. 36(1933–1934), pp. 29–95.
2. N. Jacobson, *Abstract derivation and Lie algebras*, Transactions of the American Mathematical Society, vol. 42(1937), pp. 206–224.
3. N. Jacobson, *Rational methods in the theory of Lie algebras*, Annals of Mathematics, vol. 36(1935), pp. 875–881.
4. N. Jacobson, *Restricted Lie algebras of characteristic p*, Transactions of the American Mathematical Society, vol. 50(1941), pp. 15–25.
5. W. Magnus, *Über Gruppen und zugeordnete Liesche Ringe*, Journal für die reine und angewandte Mathematik, vol. 182(1940), pp. 142–149.
6. H. Zassenhaus, *Endliche p-Gruppe und Lie-Ring mit der Characteristik p*, Abhandlungen aus dem mathematischen Seminar der Hansischen Universität, vol. 13(1939), pp. 200–206.
7. H. Zassenhaus, *Lehrbuch der Gruppentheorie*, I, Leipzig and Berlin, 1937.

The University of British Columbia.

GENERALIZED "SANDWICH" THEOREMS

By A. H. Stone and J. W. Tukey

The following theorem is well known under the self-explanatory name of the "ham sandwich theorem".[1]

Given any three sets in space, each of finite outer Lebesgue measure (m^*), *there exists a plane which bisects all three sets*, in the sense that the part of each set which lies on one side of the plane has the same outer measure as the part of the same set which lies on the other side of the plane.

The usual proof is based on the following theorem of Borsuk.[2]

If ϕ is a continuous mapping of the n-sphere S^n in Euclidean n-space R^n which is "antipodal" (i.e., diametrically opposite points of S^n map into points symmetric about the origin in R^n), then there is a point of S^n which maps into the origin of R^n.

If now p denotes a plane in R^3, let p^+ and p^- denote the two parts into which p divides R^3, and let v be the unit-vector perpendicular to p, oriented from p^- to p^+. Let A_i $(i = 1, 2, 3)$ be the given sets. The usual argument proves first, from measure-theoretic considerations, that for each v a corresponding p can be found, depending continuously on v, which bisects A_3. The correspondence $\phi(v) = [m^*(p^+ \cdot A_1) - m^*(p^- \cdot A_1), m^*(p^+ \cdot A_2) - m^*(p^- \cdot A_2)]$ is then an antipodal mapping of S^2 in R^2. The result now follows from the case $n = 2$ of the Borsuk theorem (which can, for $n = 2$, be proved readily ab initio).

Now, a fuller use of the Borsuk theorem gives an easier proof of a more general theorem. Let R be any point-set on which a Carathéodory outer measure m^* is defined. Let f be a real-valued function defined over $S^n \times R$ such that:

(1) For each $\Lambda \epsilon S^n$, $f(\Lambda, x)$ is a measurable function over R $(x \epsilon R)$, and vanishes only over a set of measure zero.

(2) For each $x \epsilon R$, $f(\Lambda, x)$ is a continuous function over S^n.

(3) For each pair of diametrically opposite points Λ and $-\Lambda$ of S^n, $f(\Lambda, x) \cdot f(-\Lambda, x) \leqq 0$ almost everywhere in R.

Write $f^+(\Lambda)$, $f^0(\Lambda)$, and $f^-(\Lambda)$ respectively for the subsets of R on which $f(\Lambda, x) > 0, = 0$, and < 0. We say "*$f^0(\Lambda)$ bisects $A \subset R$*" if $m^*(f^+(\Lambda) \cdot A) = m^*(f^-(\Lambda) \cdot A)$.

Theorem. *Given any n sets $A_1, A_2, \cdots, A_n$ in R, each of finite outer measure, there exists $\Lambda \epsilon S^n$ such that $f^0(\Lambda)$ bisects each A_i $(i = 1, 2, \cdots, n)$.*

Proof. Define a mapping ϕ of S^n in R^n by:

(4) The i-th coordinate of $\phi(\Lambda)$ is $m^*(f^+(\Lambda) \cdot A_i) - m^*(f^-(\Lambda) \cdot A_i)$.

Received November 12, 1941.

[1] Discovered by S. Ulam; we are indebted to the referee for calling this fact to our attention.

[2] Equivalent to Satz II of *Drei Sätze über die n-dimensionale euklidische Sphäre*, Fundamenta Mathematicae, vol. 20(1933), p. 177. This theorem was suggested by Ulam.

Clearly ϕ is antipodal; for, in virtue of (3) and (1), $f^+(\Lambda) = f^-(-\Lambda)$ to within sets of measure zero.

Also ϕ is continuous. For let $\{\Lambda_p\} \to \Lambda_0$ in S^n. From (2), $f^+(\Lambda_0) \subset \liminf f^+(\Lambda_p)$. Thus, since the sets $f^+(\Lambda_p)$ are measurable and A_i has finite outer measure, we have $m^*(f^+(\Lambda_0)\cdot A_i) \leqq \liminf m^*(f^+(\Lambda_p)\cdot A_i)$. On the other hand, $\limsup f^+(\Lambda_p) \subset f^+(\Lambda_0) + f^0(\Lambda_0)$, from (2). Hence $\limsup m^*(f^+(\Lambda_p)\cdot A_i) \leqq m^*(f^+(\Lambda_0)\cdot A_i) + 0$, using (1). Whence $m^*(f^+(\Lambda_p)\cdot A_i) \to m^*(f^+(\Lambda_0)\cdot A_i)$ as $p \to \infty$. A similar argument applies to f^-. This establishes the continuity of ϕ.

Hence Borsuk's theorem yields the existence of $\Lambda \epsilon S^n$ such that $\phi(\Lambda) = (0, 0, \cdots, 0)$; that is, such that $f^0(\Lambda)$ bisects each A_i.

COROLLARY. *Given $n + 1$ measurable functions $f_0, \cdots, f_n$ over R, and n sets $A_1, \cdots, A_n$ in R, of finite outer measure, then, provided that $f_0, \cdots, f_n$ are linearly independent modulo sets of measure zero (i.e., that whenever $\lambda_0 f_0(x) + \cdots + \lambda_n f_n(x) = 0$ over a subset of R of positive measure, then $\lambda_0 = \lambda_1 = \cdots = 0$), there will exist real numbers $\lambda_0, \cdots, \lambda_n$, not all zero, such that each A_i is bisected by the sets defined by $\lambda_0 f_0(x) + \cdots + \lambda_n f_n(x) < 0$ and $\lambda_0 f_0(x) + \cdots + \lambda_n f_n(x) > 0$.*

For we can take S^n as the unit sphere in R^{n+1}, and then put $f(\Lambda, x) = \lambda_0 f_0(x) + \lambda_1 f_1(x) + \cdots + \lambda_n f_n(x)$, where Λ is $(\lambda_0, \lambda_1, \cdots, \lambda_n)$.

This corollary plainly includes the "ham sandwich theorem"; we need only take $f_0 = 1$, and $f_1 = x, f_2 = y, f_3 = z$, where x, y, z are the coordinates in $R^3 = R$. It also includes such results as the following. *Any $n + 1$ sets in R^n, of finite outer measure, can be bisected by an $(n - 1)$-sphere*, in the sense that the part of each set which lies inside the sphere has the same outer measure as the part which lies outside the sphere. (A plane is here regarded as a sphere of infinite radius.) For we can take $f_0 = x_1^2 + x_2^2 + \cdots + x_n^2, f_i = x_i$ $(i = 1, 2, \cdots, n)$, and $f_{n+1} = 1$, where $x_1, x_2, \cdots, x_n$ are the coordinates in $R^n = R$.

Similarly, any five sets in the plane, of finite outer measure, can be bisected by a conic; and so on.

Thus, roughly speaking, to require that a given subset of R^n be *bisected* by one of a family of algebraic manifolds is to impose a *linear* condition. This is not true if bisection is replaced by division into other given ratios. In fact:

If α_1, α_2 are numbers such that, given any two sets A_1, A_2 in R^3, of finite measure, there exists a plane which divides A_i in the ratio $\alpha_i : 1 - \alpha_i$ $(i = 1, 2)$; then $\alpha_1 = \alpha_2 = \frac{1}{2}$.

For we can obviously suppose that $0 \leqq \alpha_1 \leqq \alpha_2 \leqq 1$; and it is easily seen that $\alpha_1 \neq 0$ and $\alpha_2 \neq 1$. Taking $A_1 = A_2$, we see that $\alpha_1 = \alpha_2$. Take A_1 to be a small "solid" sphere and A_2 to be a large concentric one; then any plane which divides A_1 in the right ratio must meet A_1 and so will approximately bisect A_2. Thus $\alpha_1 = \alpha_2 = \frac{1}{2}$.

The extension to two sets in R^n is immediate.

In R^1, however, there is a remarkable analogue.

A necessary and sufficient condition that the real numbers α_1, α_2 be such that, given any two sets A_1, A_2 in R^1, of finite outer measure, there exists an interval in R^1 whose intersection with A_i has for outer measure $\alpha_i \cdot m^*(A_i)$ $(i = 1, 2)$ is that $\alpha_1 = \alpha_2 =$ the reciprocal of an integer greater than 1.[3]

(An "interval" here is either finite or half-infinite, i.e., is of the form (a, b) or $(-\infty, b)$ or (a, ∞).)

Proof of necessity. It is easily seen, as above, that $0 < \alpha_1 = \alpha_2 < 1$. Let $\alpha_1 = \alpha_2 = \alpha$, say; then if α has not the form $1/n$, where n is an integer greater than 1, we can write

(5) $$1/(n + 1) < \alpha < 1/n, \text{ where } n \text{ is a positive integer.}$$

Let $0 < p < 1$, and denote by E_n the interval $(n - p, n + p)$. Take $A_1 = E_1 + E_3 + \cdots + E_{2n+1}$ and $A_2 = E_2 + E_4 + \cdots + E_{2n}$. Any interval I which satisfies $m(I \cdot A_1) = \alpha \cdot m(A_1)$ must (since $\alpha > 1/(n + 1)$) meet two consecutive E's of odd suffix. Hence I contains an E of even suffix; so $m(I \cdot A_2) \geqq (1/n) \cdot m(A_2) > \alpha \cdot m(A_2)$.

Proof of sufficiency. Let $\alpha_1 = \alpha_2 = 1/n$, where n is an integer greater than 1. We can take points $x_1 < x_2 < \cdots < x_{n-1} \in R^1$ such that, if I_i denotes the interval (x_i, x_{i+1}) $(i = 0, 1, \cdots, n - 1$; x_0 is written for $-\infty$, and x_n for $+\infty$), then $m^*(I_0 \cdot A_1) = m^*(I_1 \cdot A_1) = \cdots = m^*(I_{n-1} \cdot A_1) = m^*(A_1)/n$. Now, we have $m^*(I_0 \cdot A_2) + m^*(I_1 \cdot A_2) + \cdots + m^*(I_{n-1} \cdot A_2) = m^*(A_2)$. Hence either there is an i for which $m^*(I_i \cdot A_2) = m^*(A_2)/n$—in which case I_i is the required interval—or there are i, j such that $m^*(I_i \cdot A_2) < m^*(A_2)/n < m^*(I_j \cdot A_2)$. So, by an easy continuity argument, there will be an interval J (with its left-hand end-point between x_i and x_j and right-hand end-point between x_{i+1} and x_{j+1}) for which $m^*(J \cdot A_1) = m^*(A_1)/n$ and $m^*(J \cdot A_2) = m^*(A_2)/n$; and J is the required interval.

Remark. If we interpret "interval" to mean "finite interval", it is easily seen that the corresponding necessary and sufficient condition on α_1, α_2 is: $\alpha_1 = \alpha_2 =$ the reciprocal of an integer greater than 2.

The preceding result has an analogue in R^2. For let $(1/n) > \alpha > 1/(n + 1)$, and consider the following two sets: $A_2 =$ the "annulus" between two concentric similarly situated regular n-gons of sides 1 and $1 + \delta$ (where δ is small and positive) and $A_1 =$ a set of $n + 1$ equal small circles, n of which are inscribed in A_2 at its corners, and the last of which is concentric with A_2. It will readily be verified that any *circle*[4] which cuts off α times the measure of A_1 from A_1 must cut off nearly $1/n$ (at least) of A_2 in measure if δ is small. Con-

[3] This can also be deduced from a theorem of P. Lévy, *Généralisation du théorème de Rolle*, C. R. Acad. Sci., vol. 198(1934), pp. 424–425. See also H. Hopf, *Über die Sehnen ebener Kontinuen und die Schleifen geschlossener Wege*, Commentarii Mathematici Helvetici, vol. 9(1937), pp. 303–319.

[4] A half-plane (determined by a straight line) is regarded as a circle.

sequently, *a necessary condition that the real numbers* α_1, α_2 *be such that, given any two sets* A_1, A_2 *in* R^2, *of finite measure, there exists a circle which intersects* A_i *in a set of measure* $\alpha_i \cdot m(A_i)$ $(i = 1, 2)$ *is that* $\alpha_1 = \alpha_2 =$ *the reciprocal of an integer greater than* 1.

It is plausible that this condition is also sufficient, but we are unable to prove this. In the case $\alpha_1 = \alpha_2 = \frac{1}{2}$, the sufficiency follows from the "ham sandwich theorem".

THE INSTITUTE FOR ADVANCED STUDY AND PRINCETON UNIVERSITY.

THE CONTINUED FRACTION AS A SEQUENCE OF LINEAR TRANSFORMATIONS

BY J. FINDLAY PAYDON AND H. S. WALL

1. **Introduction.** This paper contains a development of properties of the continued fraction

$$\frac{1}{1} + \frac{a_2}{1} + \frac{a_3}{1} + \frac{a_4}{1} + \cdots \tag{1.1}$$

in which the elements $a_2, a_3, a_4, \cdots$ are complex numbers. The central idea may be described as follows. With the continued fraction (1.1) we associate a sequence of linear transformations[1]

$$a_1(v) = v, \qquad a_k(v) = \frac{1}{1 + a_k v} \qquad (k = 2, 3, 4, \cdots). \tag{1.2}$$

Then the product of the first n of these is

$$a_1 a_2 \cdots a_n(v) = \frac{1}{1} + \frac{a_2}{1} + \cdots + \frac{a_{n-1}}{1} + \frac{a_n v}{1} = \frac{a_n A_{n-2} v + A_{n-1}}{a_n B_{n-2} v + B_{n-1}}, \tag{1.3}$$

where A_k and B_k are the k-th numerator and denominator of (1.1), i.e., $A_{-1} = 1$, $B_{-1} = 0$, $A_0 = 0$, $B_0 = 1$, $A_k = A_{k-1} + a_k A_{k-2}$, $B_k = B_{k-1} + a_k B_{k-2}$ ($k = 1, 2, 3, \cdots$; $a_1 = 1$). Corresponding to an arbitrary set V of points in the complex z-plane, called a *value region*, we determine a set $\mathfrak{A}$ of points, called the *element region* (corresponding to V), by the condition that a is in $\mathfrak{A}$ if and only if the transformation $w = a(v) = 1/(1 + av)$ transforms V into a subset of itself, i.e., $a(V) \subset V$. It is at once evident that if $a_2, a_3, a_4, \cdots$ are in $\mathfrak{A}$, and

$$a_1 a_2 \cdots a_n(V) = V^{(n)} \qquad (n = 1, 2, 3, \cdots),$$

then

$$V = V^{(1)} \supset V^{(2)} \supset V^{(3)} \supset \cdots.$$

Hence, if V is a bounded closed set so that $V^{(1)}, V^{(2)}, V^{(3)}, \cdots$ are closed, then there are just two cases, namely:

Case I. The sets $V^{(n)}$ $(n = 1, 2, 3, \cdots)$ have one and only one point, v_0, in common.

Case II. The sets $V^{(n)}$ $(n = 1, 2, 3, \cdots)$ have two or more points in common.

In Case I, we have, *uniformly* for all v in V:

$$\lim_{n=\infty} a_1 a_2 \cdots a_n(v) = v_0.$$

Received November 12, 1941. Part of this paper was presented to the American Mathematical Society by J. Findlay Paydon, September 2, 1941, under the title *Convergence regions and value regions for continued fractions*, and part by J. Findlay Paydon and H. S. Wall, December 29, 1941, under the title *An extension of the Stieltjes continued fraction theory.*

[1] It is convenient to use the symbol a_k in two senses. The subscript 1 will be reserved for the identity transformation.

Hence it follows that if V contains 0 or 1, the continued fraction converges to v_0 inasmuch as $A_{n-1}/B_{n-1} = a_1 a_2 \cdots a_n(0)$ and $A_n/B_n = a_1 a_2 \cdots a_n(1)$. In Case II the continued fraction is in general divergent, and in Case I the continued fraction may diverge if V does not contain 0 or 1.[2]

We take for value region $V = V_c$ the circle $|z - c| = |c|$ and its interior, where $\Re(c) > \frac{1}{2}$, and determine the corresponding element region $\mathfrak{A} = \mathfrak{A}_c$ to be a parabola and its interior, the parabola having its focus at the origin and not having the point $-\frac{1}{4}$ on the interior. If a is an arbitrary point not in the interval $[\Re(a) \leqq -\frac{1}{4}, \Im(a) = 0]$, then c can be found such that $\Re(c) > \frac{1}{2}$, and such that a lies within $\mathfrak{A}_c$. We then show in particular that (1.1) converges if $a_2, a_3, a_4, \cdots$ lie in any bounded closed region within $\mathfrak{A}_c$.

In §2 we determine the element region $\mathfrak{A}_c$ corresponding to the value region V_c defined above. In §3 we consider in detail the case where c is real, and arrive at new proofs of the "parabola theorem" and the "parabola-circle theorem" of Scott and Wall.[3] In §4 there is a determination of the value region in case the element region is $|z| \leqq r \leqq \frac{1}{4}$ (which is a subregion of $\mathfrak{A}_1$). The main result is contained in §5, and may be described as an extension of the Stieltjes[4] convergence theorem.

If $a_2, a_3, a_4, \cdots$ are in the horizontal strip $-\frac{1}{2}h \leqq y \leqq +\frac{1}{2}h$ in the plane of $z = x + iy$, where $0 < h \leqq 1$, then the continued fraction

$$\frac{1}{1} + \frac{a_2^2 t}{1} + \frac{a_3^2 t}{1} + \frac{a_4^2 t}{1} + \cdots \tag{1.4}$$

converges uniformly over any bounded closed region lying entirely within the cardioid

$$\rho = \frac{1}{2h^2}(1 + \cos\theta), \qquad t = \rho e^{i\theta}, \tag{1.5}$$

provided the series $\sum |b_n|$ diverges,[5] where $b_1 = 1$, $b_{n+1} = 1/b_n a_{n+1}^2$ $(n = 1, 2, 3, \cdots)$. If the series $\sum |b_n|$ converges, then the sequences of even and odd approximants converge to separate limits which are meromorphic functions of $1/t$.

The Stieltjes convergence theorem appears as the limiting case $h = 0$. The function represented by a continued fraction may have a singularity at any assigned point upon the cardioid, and therefore the result is in a certain sense the best. The concluding section of the paper contains a discussion of a class of continued fractions with elements in the unit circle.

[2] For example, if V consists of the single point t $(t \neq 0, 1)$, then $\mathfrak{A}$ contains the single point $a = (1 - t)/t^2$, and the continued fraction diverges if a is real and less than $-1/4$.

[3] W. T. Scott and H. S. Wall, (1) *A convergence theorem for continued fractions*, Transactions of the American Mathematical Society, vol. 47(1940), pp. 155–172; and (2) *Value regions for continued fractions*, Bulletin of the American Mathematical Society, vol. 47 (1941), pp. 580–585.

[4] T. J. Stieltjes, *Recherches sur les fractions continues*, *Oeuvres*, vol. II, 1918, pp. 402–566.

[5] This series is to be counted as divergent if some a_n is 0.

2. $(V_c, \mathfrak{A}_c)$. Let $c = r + is$, where r, s are real and $\frac{1}{2} < r \leqq 1$, and denote by V_c the circle $|z - c| = |c|$ and its interior. The element region $\mathfrak{A}_c$ corresponding to V_c as value region consists of all points a such that the transformation $a(v) \coloneqq 1/(1 + av)$ carries V_c into a circular region in V_c. Put $a = x + iy$, $w = X + iY$. Then the transformation $w = a(v)$ takes $|v - c| = |c|$ into the circle

$$(X - \alpha)^2 + (Y - \beta)^2 = \gamma^2, \tag{2.1}$$

where

$$\alpha = \frac{1 + rx - sy}{1 + 2(rx - sy)}, \quad \beta = -\frac{sx + ry}{1 + 2(rx - sy)}, \quad \gamma^2 = \frac{(rx - sy)^2 + (sx + ry)^2}{[1 + 2(rx - sy)]^2}.$$

This circle evidently passes through 1, and consequently will lie in the circle $|z - c| = |c|$ if and only if its center lies within or upon the ellipse $J: |z - c| + |z - 1| = |c|$, which is the locus of centers of circles through 1 tangent to the circle $|z - c| = |c|$. The equation of J in rectangular coordinates X, Y is:

$$(1 - 2r - s^2)X^2 + 2s(r - 1)XY - r^2Y^2 + (2r^2 + 2s^2 + r - 1)X$$
$$+ sY + (1 - 4r^2 - 4s^2)/4 = 0.$$

Hence $a = x + iy$ must satisfy the inequality obtained by putting $X = \alpha$, $Y = \beta$ and replacing "$=$" by "$\geqq$" in the last equation. On doing this and simplifying we obtain the inequality:

$$\left[\frac{2rsx - (s^2 - r^2)y}{r^2 + s^2}\right]^2 \leqq \frac{1 - 2r}{r^2 + s^2}\left[\frac{2rsy + (s^2 - r^2)x}{r^2 + s^2}\right] + \left[\frac{1 - 2r}{2(r^2 + s^2)}\right]^2. \tag{2.2}$$

When $a = x + iy$ satisfies this inequality, it is easy to see that the interior of (2.1) lies in the interior of V_c. *Hence we have proved that $\mathfrak{A}_c$ consists of the points $a = x + iy$ which satisfy* (2.2).

Inasmuch as interior points of V_c are mapped into interior points of (2.1), and since 1 is interior to V_c, it follows that when $a_2, a_3, a_4, \cdots$ are in the parabolic region (2.2), then all the approximants $a_1 a_2 \cdots a_n(1)$ of (1.1) are in the *interior* of V_c.

In order to throw (2.2) into a more convenient form, put

$$\begin{aligned} x' &= x\cos\phi + y\sin\phi \\ y' &= -x\sin\phi + y\cos\phi \end{aligned} \qquad \left(\cos\phi = \frac{r^2 - s^2}{r^2 + s^2}, \quad \sin\phi = \frac{-2rs}{r^2 + s^2}\right),$$

and (2.2) becomes $y'^2 \leqq P\left(x' + \dfrac{P}{4}\right)$, where $P = (2r - 1)/(r^2 + s^2)$. It will be observed that $\phi = -2 \arg c$. The polar equation of the parabola bounding $\mathfrak{A}_c$ is

$$\rho = \frac{P}{2[1 - \cos(\theta - \phi)]}, \qquad z = \rho e^{i\theta}. \tag{2.3}$$

This parabola never contains the point $-\frac{1}{4}$ in its interior; it passes through $-\frac{1}{4}$, and therefore has maximum extent when $r = 1$, in which case (2.3) may be written:

$$\rho = \frac{\frac{1}{2}\cos^2 \frac{1}{2}\phi}{1 - \cos(\theta - \phi)}. \tag{2.4}$$

We shall summarize the main results of this section as

THEOREM 2.1. *If the elements $a_2, a_3, a_4, \cdots$ of* (1.1) *lie within or upon the parabola*

$$| z | - \Re(z)\cos\phi - \Im(z)\sin\phi = \tfrac{1}{2}P, \tag{2.5}$$

where $-\pi < \phi < \pi$, $P = (2r-1)/(r^2+s^2)$, $\frac{1}{2} < r \leqq 1$, $s/r = -\tan\frac{1}{2}\phi$, then all its approximants lie within the circle $| z - c | = | c |$ where $c = r + is$.

3. $(V_r, \mathfrak{A}_r)$, $\frac{1}{2} < r \leqq 1$. When $c = r$, the preceding theorem becomes:

THEOREM 3.1. *The element region $\mathfrak{A}_r$ corresponding to the value region V_r: $| z - r | \leqq r$ $(\frac{1}{2} < r \leqq 1)$ is the parabolic region $y^2 \leqq h(x + h/4)$, $z = x + iy$, where $h = (2r-1)/r^2$. If $a_2, a_3, a_4, \cdots$ are in $\mathfrak{A}_r$, then all the approximants of the continued fraction* (1.1) *lie in the interior of V_r.*

We shall now obtain conditions on $a_2, a_3, a_4, \cdots$ which are necessary and sufficient for Case I (cf. §1). Since $\mathfrak{A}_1 \supset \mathfrak{A}_r$ for $\frac{1}{2} < r \leqq 1$, we may as well assume that $r = 1$. Let $a_1a_2 \cdots a_n(V_1) = V^{(n)}$; denote by $K^{(n)}$ the circle bounding $V^{(n)}$; and let $R^{(n)}$ be the radius of $K^{(n)}$. Then, on applying (1.3) to the circle $K^{(1)}$, we find for $R^{(n)}$ the value

$$R^{(n)} = \frac{| a_2a_3 \cdots a_n |}{| B_n |^2 - | a_nB_{n-2} |^2} \qquad (n = 1, 2, 3, \cdots). \tag{3.1}$$

If some a_n is 0, let a_k be the first which is 0. Then, inasmuch as $| B_k |^2 - | a_kB_{k-2} |^2 = | B_k |^2 = | B_{k-1} |^2$, and $B_{k-1} \neq 0$ because of the relation $| A_{k-1}B_{k-2} - A_{k-2}B_{k-1} | = | a_2a_3 \cdots a_{k-1} | \neq 0$ and the fact that A_{k-1}/B_{k-1} is finite, it follows that $R^{(k)} = 0$. Hence, also, $R^{(n)} = 0$ for $n = k+1, k+2, \cdots$, so that $V^{(1)}$, $V^{(2)}, \cdots$ have one and only one point in common, and we therefore have Case I.

We now use the relations

$$B_{n+2} = (1 + a_{n+1} + a_{n+2})B_n - a_na_{n+1}B_{n-2} \quad (n = 2, 3, 4, \cdots) \tag{3.2}$$

to obtain an inequality for $R^{(n)}$ in case $a_n \neq 0$ $(n = 2, 3, 4, \cdots)$. Since the a_n's are in the parabolic region $\mathfrak{A}_1$ so that $| a_n | = \Re(a_n) + h_n/2$, where $0 \leqq h_n \leqq 1$ $(n = 2, 3, 4, \cdots)$, we readily find that $| 1 + a_2 | > | a_2 |$, $| 1 + a_2 + a_3 | > | a_3 |$, $| 1 + a_n + a_{n+1} | \geqq | a_n | + | a_{n+1} |$ $(n = 3, 4, 5, \cdots)$. Hence if $b_1 = 1$, $b_{n+1} = 1/b_na_{n+1}$ $(n = 2, 3, 4, \cdots)$, we have:

$$\begin{aligned} &| 1 + b_1b_2 | > 1, \qquad | 1 + b_3 + b_2b_3 | > 1, \\ &| b_{n-1}b_nb_{n+1} + b_{n-1} + b_{n+1} | \geqq | b_{n+1} | + | b_{n-1} | \qquad (n = 3, 4, 5, \cdots). \end{aligned} \tag{3.3}$$

If we put $Q_n = b_1 b_2 \cdots b_n B_n$, the first of these may be written

$$|Q_2| \geqq 1 + k|b_2|, \qquad |Q_3| \geqq 1 + k|b_3|, \tag{3.4}$$

where $k > 0$. On making the same substitutions in (3.2) we get

$$b_n Q_{n+2} = (b_n b_{n+1} b_{n+2} + b_n + b_{n+2})Q_n - b_{n+2}Q_{n-2} \qquad (n = 2, 3, 4, \cdots),$$

from which we obtain by means of (3.3) the inequalities:

$$|Q_{n+2}| - |Q_n| \geqq \left|\frac{b_{n+2}}{b_n}\right| \{|Q_n| - |Q_{n-2}|\} \qquad (n = 2, 3, 4, \cdots),$$

and consequently:

$$|Q_{n+2}| - |Q_n| \geqq k|b_{n+2}| \qquad (n = 2, 3, 4, \cdots). \tag{3.5}$$

By (3.4) this holds also if $n = 0, 1$.

In terms of the Q_n's and b_n's our expression for $R^{(n)}$ in (3.1) becomes:

$$R^{(n)} = \frac{|b_n|}{|Q_n| - |Q_{n-2}|} \cdot \frac{1}{|Q_n| + |Q_{n-2}|},$$

and therefore, by (3.5) and (3.4),

$$R^{(2n)} < \frac{1}{2k\left(1 + k\sum_{i=1}^{n-1} |b_{2i}|\right)}, \qquad R^{(2n+1)} < \frac{1}{2k\left(1 + k\sum_{i=1}^{n-1} |b_{2i+1}|\right)}. \tag{3.6}$$

These inequalities together with the inequality $R^{(n)} \leqq R^{(n-1)}$ show that if the series $\sum |b_n|$ is divergent then $R^{(n)} \to 0$ as $n \to \infty$. On the other hand, when the series $\sum |b_n|$ converges,[6] we know that the sequences of even and odd approximants converge to separate limits L_0, L_1, and therefore, inasmuch as

$$R^{(n)} \geqq \left|\frac{A_n}{B_n} - \frac{A_{n-1}}{B_{n-1}}\right|,$$

we conclude that $R^{(n)}$ does not tend to 0 as n tends to ∞ in this case. Since $R^{(2)} \leqq \frac{1}{2}$ we must have $|L_0 - L_1| \leqq 1$. We have completed the proof of the following theorem.

THEOREM 3.2. *If V_r, $\mathfrak{A}_r$ are as in Theorem* 3.1, *the following conditions are necessary and sufficient for Case* I (cf. §1): (i) *some a_n is* 0; *or* (ii) $a_n \neq 0$ ($n = 2, 3, 4, \cdots$), *and the series $\sum |b_n|$ diverges, where $b_1 = 1$, $b_{n+1} = 1/b_n a_{n+1}$* ($n = 1, 2, 3, \cdots$). *In Case* I *the continued fraction* (1.1) *converges and its value is in V_r. In Case* II (cf. §1), *the continued fraction diverges by oscillation, the sequence of even approximants having a limit L_0 and the sequence of odd approximants a limit L_1. We have: $L_0 \neq L_1$; L_0 and L_1 are in V_r; $|L_0 - L_1| \leqq 1$.*

[6] O. Perron, *Die Lehre von den Kettenbrüchen*, second edition, Leipzig and Berlin, 1929, pp. 235–236.

Scott and Wall[7] showed that, if w is a value $\neq 0$ upon the circle $K^{(1)}$, $|z-1| = 1$, then there is a continued fraction of the form

$$\frac{1}{1} + \frac{a}{1} + \frac{\bar{a}}{1} + \frac{a}{1} + \frac{\bar{a}}{1} + \cdots, \tag{3.7}$$

where a lies on the boundary of $\mathfrak{A}_1$ and $\bar{a}$ is the complex conjugate of a, which converges to the value w. *We shall now prove that* (3.7) *is the only continued fraction with elements in* $\mathfrak{A}_1$ *whose value is* w. More generally, we have the following theorem.

THEOREM 3.3. *If a value region V is a region whose boundary is a circle K passing through the point* 0 *and containing* 1 *on the interior with corresponding element region* $\mathfrak{A}$, *and if there is a continued fraction* (1.1) *with elements in* $\mathfrak{A}$ *which converges to a value w upon K, then there is only one such continued fraction.*

Proof. Since 1 is interior to V, all the approximants $a_1a_2 \cdots a_n(1)$ of (1.1) are interior to V so that no approximant can equal w. Let $a_1a_2 \cdots a_n(K) = K^{(n)}$. If (1.1) converges to the value w, then $K^{(n)}$ must be tangent to K at w; and since $a_1a_2 \cdots a_{n-1}(1) = a_1a_2 \cdots a_n(0)$ lies on $K^{(n)}$, it follows that a_n is uniquely determined in terms of $a_2, a_3, \cdots, a_{n-1}$. The theorem now follows by mathematical induction.

4. **The Worpitzky circle.** Continued fractions of the form (1.1) whose elements lie in the neighborhood of the origin are of particular importance in the applications to function theory. Worpitzky showed that (1.1) converges when the a_n's are in the circular region $|z| \leqq \frac{1}{4}$. Since this region is contained in $\mathfrak{A}_1$, where $\mathfrak{A}_1$ is the parabolic region of §3, the Worpitzky result is included in Theorem 3.2. Moreover, from Theorem 3.2 it follows that the value of (1.1) lies in V_1 when its elements are in the Worpitzky circle. We shall now obtain a better estimate for the value of (1.1) in this case. The result is as follows.

THEOREM 4.1. *If $a_2, a_3, a_4, \cdots$ lie within or upon the circle*

$$|z| = (2r-1)/4r^2 \qquad (\tfrac{1}{2} < r \leqq 1), \tag{4.1}$$

then all the approximants of the continued fraction (1.1) *lie within the circle*

$$\left| z - \frac{4r^2}{4r-1} \right| = \frac{2r(2r-1)}{4r-1}. \tag{4.2}$$

There is at least one continued fraction with elements within or upon (4.1) *whose value is any preassigned number within or upon* (4.2); *and if w is any number upon* (4.2) *there is one and only one continued fraction with elements within or upon* (4.1) *whose value is w, namely*:

$$\frac{1}{1} + \frac{a_2}{1} + \frac{(1-2r)/4r^2}{1} + \frac{(1-2r)/4r^2}{1} + \cdots, \tag{4.3}$$

[7] Scott and Wall, footnote 3, (1), p. 166.

where, if

$$w = \frac{4r^2 + 2r(2r-1)e^{i\phi}}{4r-1} \qquad (0 \leqq \phi < 2\pi), \tag{4.4}$$

then

$$a_2 = -\frac{(2r-1)}{4r^2}\cdot\frac{\cos\phi + (8r^2-4r)(1+\cos\phi) + i(4r-1)\sin\phi}{1+(8r^2-4r)(1+\cos\phi)}. \tag{4.5}$$

Proof. If a_2 is in the circle (4.1), the value of the continued fraction (4.3) is $w = 1/(1 + 2ra_2)$; and as a_2 ranges over the circle (4.1) and its interior, w ranges over the circle (4.2) and its interior. Hence there is at least one continued fraction with elements in (4.1) which takes on any preassigned value in (4.2).

To prove the first part of the theorem it suffices to show that if v is any value *within* the circle (4.2) then $w = 1/(1 + a_2v)$ is also within this circle for all values of a_2 in the circle (4.1). Now if v is fixed, we may consider $w = 1/(1 + a_2v)$ as a transformation of a_2 into w; and we find that it carries the circle (4.1) into the circle with center and radius:

$$C = \frac{16r^4}{16r^4 - (2r-1)^2|v|^2}, \qquad R = \frac{4r^2(2r-1)|v|}{16r^4 - (2r-1)^2|v|^2}.$$

Inasmuch as this circle lies within the circle (4.2) when $|v| < 2r$, it follows that every approximant of (1.1) lies *within* (4.2).

Let w be any value upon the circle (4.2), expressed in the form (4.4). We may write $w = 1/(1 + a_2v)$, where v is the value of a convergent continued fraction with elements in (4.1). We then have:

$$\begin{aligned} a_2v &= \frac{1-w}{w} \\ &= \frac{1-2r}{2r}\cdot\frac{\cos\phi + (8r^2-4r)(1+\cos\phi) + i(4r-1)\sin\phi}{1+(8r^2-4r)(1+\cos\phi)}, \end{aligned} \tag{4.6}$$

and thus it follows that $|a_2v| = (2r-1)/2r$, or $|a_2| = [(2r-1)/4r^2]\cdot(2r/|v|)$. Consequently, inasmuch as $|v| \leqq 2r$, we conclude that $|a_2| \geqq (2r-1)/4r^2$. But $|a_2| \leqq (2r-1)/4r^2$, by hypothesis. Therefore,

$$|a_2| = (2r-1)/4r^2, \qquad |v| = 2r, \qquad v = 2r,$$

where the last equation follows from the fact that v is in the circle (4.2). On putting $v = 2r$ in (4.6) we then find that a_2 must have the value (4.5). Now, starting with $v = 2r$ as the value upon the circle (4.2) to be attained, we find in the same way that a_3 must be given by the expression in the right member of (4.6) divided by $2r$, but with ϕ now equal to 0. This gives for a_3 the value $(1-2r)/4r^2$, and on repeating the argument we find that $a_4, a_5, \cdots$ must all have this same value. We have completed the proof of Theorem 4.1.

A reexamination of the preceding proof will show that a stronger theorem holds for the case $r = 1$, namely:

THEOREM 4.2. *If* $|a_2| \leqq \frac{1}{4}$, *and* $a_3, a_4, \cdots$ *are in the parabola* $|z| - R(z) = \frac{1}{2}$, *then all the approximants of the continued fraction* (1.1) *lie within the circle* $|z - (\frac{4}{3})| = \frac{2}{3}$. *Among all the continued fractions* (1.1) *with* $|a_2| \leqq \frac{1}{4}$ *and* $a_3, a_4, \cdots$ *in the parabola, there is one and only one which takes on a prescribed value upon the circle.*

5. **The main theorem.** If $a_2, a_3, \cdots$ are in the parabolic region $\mathfrak{A}_r$ of §3 $(\frac{1}{2} < r \leqq 1)$ and $t = |t|e^{i\phi}$ $(-\pi < \phi < +\pi)$, then $a_2t, a_3t, \cdots$ are in the parabolic region $\mathfrak{A}_c$ of §2 for an appropriate value of c, if and only if

$$|t| \leqq \frac{1}{2h^2}(1 + \cos\phi), \qquad h^2 = (2r - 1)/r^2.$$

Hence it follows that if $a_2, a_3, \cdots$ are in the parabola

$$y^2 = h^2(x + h^2/4) \qquad (0 < h \leqq 1),$$

and t is in the portion of the cardioid $\rho = (1 + \cos\theta)/2h^2$ inside the sector $-\pi + \alpha \leqq \theta \leqq +\pi - \alpha$, where α is an arbitrarily small positive number, then all the approximants of the continued fraction

$$\frac{1}{1} + \frac{a_2t}{1} + \frac{a_3t}{1} + \frac{a_4t}{1} + \cdots \tag{5.1}$$

lie in a bounded region consisting of the interiors of two circles depending upon α. If G is any closed region entirely within the cardioid, then $\alpha > 0$ can be found so that G will lie in the specified portion of the cardioid. *Consequently, the approximants of* (5.1) *are uniformly bounded over* G.

Now, when $0 < t < 1/h^2$ we know that $a_2t, a_3t, \cdots$ lie in the parabolic region $\mathfrak{A}_1$, and consequently, by the "parabola theorem", (5.1) converges for these real values of t provided the series $\sum |b_n|$ diverges, where $b_1 = 1$, $b_{n+1} = 1/a_{n+1}b_n$ $(n = 1, 2, 3, \cdots)$, whereas the sequences of even and odd approximants converge to separate limits for these values of t if $\sum |b_n|$ converges. We may therefore apply the *Stieltjes-Vitali theorem* and conclude that (5.1) converges uniformly over G when $\sum |b_n|$ diverges, whereas the sequences of even and odd approximants converge uniformly over G to separate limits when $\sum |b_n|$ converges. The limits are in all cases analytic functions of t over G. Of course, it is well known that when $\sum |b_n|$ converges the sequences of even and odd approximants converge over the whole plane with the exception of isolated values of t, and the limit functions are meromorphic functions of $1/t$.

We shall state our result as

THEOREM 5.1. *Let* $0 < h \leqq 1$ *and let* $a_2, a_3, \cdots$ *be any numbers lying within or upon the parabola*

$$|z| - \Re(z) = \tfrac{1}{2}h^2. \tag{5.2}$$

Let G be any closed region entirely within the cardioid (1.5). *Then the continued fraction* (5.1) *converges uniformly over G if the series* $\sum |b_n|$ *diverges, where* $b_1 = 1$, $b_{n+1} = 1/b_n a_{n+1}$ $(n = 1, 2, 3, \cdots)$, *while the sequences of even and odd approximants converge uniformly over G to separate limits if* $\sum |b_n|$ *converges.*

It will be noted that when a_n is in the strip $-\frac{1}{2}h \leqq y \leqq +\frac{1}{2}h$ $(z = x + iy)$, then a_n^2 is in the parabola (5.2), and hence our theorem can be stated in terms of the continued fraction (1.4) as was done in §1.

The continued fraction $1/1 + at/1 + at/1 + at/1 + \cdots$ in which a is any value upon the parabola (5.2) represents a function $f(t)$ having a branch point at $t = -1/4a$, which is a point upon the cardioid (1.5). Hence our theorem does not hold if the cardioid is replaced by a curve enclosing points outside the cardioid.

If $a_2, a_3, \cdots$ are real and positive and G is an arbitrary bounded closed region at a positive distance from the negative half of the real axis, then, by choosing $h > 0$ sufficiently small, the cardioid (1.5) can be made large enough to contain G on its interior, and at the same time the a_n's will remain in the parabola (5.2). Hence the continued fraction converges uniformly over G if $\sum b_n$ diverges, and the sequences of even and odd approximants converge uniformly over G if $\sum b_n$ converges. Thus we obtain the Stieltjes convergence theorem as a limiting case of Theorem 5.1.

As a corollary to Theorem 5.1 we have at once the following theorem about the continued fraction (1.1) and the element region $\mathfrak{A}_c$ of §2.

THEOREM 5.2. *If* $c = r + is$ $(\frac{1}{2} < r < 1)$ *and* $a_2, a_3, \cdots$ *are in* $\mathfrak{A}_c$ (*cf.* §2), *then the continued fraction* (1.1) *converges if and only if the series* $\sum |b_n|$ *diverges. In particular,* (1.1) *converges if the* a_n*'s are in any bounded portion of* $\mathfrak{A}_c$.

If a is any point of the plane which is not in the interval $[\Re(a) \leqq -\frac{1}{4}, \Im(a) = 0]$ c can be so chosen that $\mathfrak{A}_c$ contains a on the interior. Hence we have this theorem.

THEOREM 5.3.[8] *If a is not in the interval* $[\Re(a) \leqq -\frac{1}{4}, \Im(a) = 0]$, *then a domain D of arbitrarily large (finite) area can be found containing a on the interior such that* (1.1) *converges if the* a_n*'s are in D.*

The following theorem of E. B. Van Vleck[9] is easy to establish with the aid of the preceding ideas.

THEOREM 5.4. *If* $a_1, a_2, a_3, \cdots$ *is a sequence of numbers having a finite limit* $a \neq 0$, *and G is any bounded closed region containing no point of the ray*

[8] Otto Szász established the existence of a circular domain D with center a such that (1.1) converges if the a_n's are in D. Cf. Perron, footnote 6, p. 282.

[9] E. B. Van Vleck, *On the convergence of algebraic continued fractions whose coefficients have limiting values*, Transactions of the American Mathematical Society, vol. 5(1904), pp. 253–262. Cf. Perron, footnote 6, p. 288.

from $-1/4a$ *to* ∞ *in the direction of the vector* $-1/4a$, *then there exists an* N *such that if* $n \geqq N$ *the continued fraction*

$$\frac{1}{1} + \frac{a_{n+1}t}{1} + \frac{a_{n+2}t}{1} + \cdots \tag{5.3}$$

converges uniformly over the region G.

Proof. The region aG is a bounded closed region containing no point of the interval $[\Re(a) \leqq -\frac{1}{4}, \Im(a) = 0]$. Let $k > 1$ and $h > 0$ be so chosen that aG will lie on the interior of the region bounded by the cardioid (1.5) together with the circle $| z + 1/8k | = 1/8k$. Then if t is in G so that at is in aG and N is sufficiently large, we shall have a_n/a in the parabola (5.2), and at the same time $| a_n/a | < k$, for $n > N$. It follows that if $n \geqq N$ the approximants of the continued fraction

$$\frac{1}{1} + \frac{(a_{n+1}/a)at}{1} + \frac{(a_{n+2}/a)at}{1} + \cdots$$

are uniformly bounded for t in G. In fact, if at is in the circle $| z + 1/8k | = 1/8k$, then $| (a_n/a)at | < k(1/4k) = 1/4$ for $n > N$, so that the approximants are all in the circle $| z - 1 | = 1$; while if at is in aG but outside the circle $| z + 1/8k | = 1/8k$, the approximants are in one or the other of two fixed circles, as in the proof of Theorem 5.1. Therefore the continued fraction (5.3) converges uniformly over G if $n \geqq N$ inasmuch as the series $\sum | b_n |$ constructed for (5.3) is evidently divergent.

6. A class of continued fractions with elements in the unit circle.

In 1901, E. B. Van Vleck[10] proved a theorem which may be stated in the following form.

THEOREM 6.1. *If* $g_1, g_2, \cdots$ *are real numbers such that* $0 \leqq g_n < 1$ $(n = 1, 2, \cdots)$ *and* $x_1, x_2, \cdots$ *are independent complex variables, then the continued fraction*

$$\frac{1}{1} + \frac{g_1 x_1}{1} + \frac{(1 - g_1)g_2 x_2}{1} + \frac{(1 - g_2)g_3 x_3}{1} + \cdots \tag{6.1}$$

converges uniformly for $| x_n | \leqq 1$ $(n = 1, 2, \cdots)$ *provided the series*

$$1 + \sum_{n=1}^{\infty} \frac{g_1 g_2 \cdots g_n}{(1 - g_1)(1 - g_2) \cdots (1 - g_n)} \tag{6.2}$$

converges. The sum of the series (6.2) *is the value of the continued fraction for* $x_1 = x_2 = \cdots = -1$, *and is the least upper bound of the absolute value of the continued fraction for* $| x_n | \leqq 1$ $(n = 1, 2, \cdots)$.

[10] E. B. Van Vleck, *On the convergence and character of the continued fraction* $a_1z/1 + a_2z/1 + a_3z/1 + \cdots$, Transactions of the American Mathematical Society, vol. 2(1901), pp. 476–483.

Closely related to this theorem of E. B. Van Vleck is the following theorem which includes the *Pringsheim criteria.*[11]

THEOREM 6.2. *If* $0 \leqq g_n < 1$ *or* $0 < g_n \leqq 1$ $(n = 1, 2, \cdots)$, *then the continued fraction*

$$\frac{g_1}{1} + \frac{(1 - g_1)g_2 x_2}{1} + \frac{(1 - g_2)g_3 x_3}{1} + \cdots \tag{6.3}$$

converges uniformly for $|x_n| \leqq 1$ $(n = 2, 3, \cdots)$, *its value for* $x_2 = x_3 = \cdots = -1$ *is* $1 - (1/S)$ *where* S *is the sum of the series* (6.2), *and the absolute value of the continued fraction does not exceed* $1 - (1/S)$ *for* $|x_n| \leqq 1$ $(n = 2, 3, \cdots)$.

It is easy to see that Theorem 6.1 is a consequence of Theorem 6.2. For, if we multiply (6.3) by x_1, add 1, and take the reciprocal of the resulting continued fraction, we obtain (6.1); and the uniform convergence of (6.1) follows from the fact that $1 - (1/S) < 1$ if (6.2) converges.

Our object here is to give a simple proof of Theorem 6.1, to show, conversely, that Theorem 6.2 is a consequence of Theorem 6.1, and, finally, using the method of linear transformations, to show that (6.1) always converges for $|x_n| \leqq 1$ $(n = 1, 2, \cdots)$, except possibly for $x_1 = x_2 = \cdots = -1$.

(i) *Proof of Theorem* 6.1. If $x_1 = x_2 = \cdots = -1$ and $g_n = t_n/(1 + t_n)$ $(n = 1, 2, \cdots)$, then (6.1) is equivalent to the continued fraction

$$\frac{1}{1} - \frac{t_1}{1 + t_1} - \frac{t_2}{1 + t_2} - \cdots, \tag{6.4}$$

whose n-th numerator and denominator are $G_n = 1 + t_1 + t_1t_2 + \cdots + t_1t_2 \cdots t_{n-1}$ and $H_n = 1$, respectively. Thus we see that (6.4) is equivalent to the series $1 + \sum t_1t_2 \cdots t_n$, which is the same as the series (6.2). Hence it follows that (6.1) converges for $x_1 = x_2 = \cdots = -1$ if and only if the series (6.2) converges, and the series and continued fraction are equal when the x's have this value.

We shall show next that if $|x_n| \leqq 1$ $(n = 1, 2, \cdots)$, then (6.1) can be written in the form

$$\frac{1}{1} - \frac{r_1}{1 + r_1} - \frac{r_2}{1 + r_2} - \cdots, \tag{6.5}$$

where

$$|r_n| \leqq t_n \qquad (n = 1, 2, \cdots); \tag{6.6}$$

and since (6.5) is equivalent to the series $1 + \sum r_1r_2 \cdots r_n$, the truth of the theorem will be evident.

[11] Cf. Perron, footnote 6, pp. 254–264. Cf. also Scott and Wall, footnote 4, (1), pp. 158–160. Following Perron, Scott and Wall erroneously ascribe Theorem 3.2, p. 159, to Van Vleck.

Let $r_1, r_2, \cdots$ be defined recursively by the formula

$$r_n = \frac{-k_{n+1}x_n(1 + r_{n-1})}{1 + k_{n+1}x_n(1 + r_{n-1})} \qquad (n = 1, 2, \cdots), \tag{6.7}$$

where $r_0 = 0$, and $k_{n+1} = t_n/(1 + t_{n-1})(1 + t_n)$ $(n = 1, 2, \cdots; t_0 = 0)$. Then, $|r_1| = |k_2x_1/(1 + k_2x_1)| \leqq k_2/(1 - k_2) = t_1$, and, by mathematical induction, if $|r_{n-1}| \leqq t_{n-1}$ then $|r_n| = |k_{n+1}x_n(1 + r_{n-1})/\{1 + k_{n+1}x_n(1 + r_{n-1})\}| \leqq k_{n+1}(1 + t_{n-1})/\{1 - k_{n+1}(1 + t_{n-1})\} = t_n$, so that (6.6) holds. Now if B_n is the n-th denominator of (6.1) then we have:

$$r_n = \frac{-k_{n+1}x_nB_{n-1}}{B_{n+1}}, \quad \text{and} \quad B_{n+1} \neq 0 \qquad (n = 1, 2, \cdots). \tag{6.8}$$

For, $B_0 = B_1 = 1$, $B_2 = 1 + k_2x_1 \neq 0$, and the formula is evidently true when $n = 1$. Assuming that (6.8) holds for $n \leqq p$, it follows at once from (6.7) and (6.6) that (6.8) holds for $n = p + 1$, and hence for all n. On substituting the values of the r's from (6.8) into (6.5) the latter may be readily transformed into (6.1), and the proof is complete.

(ii) *Under either of the hypotheses of Theorem* 6.2, *the continued fraction* (6.3) *can be written in the form*

$$g_1\left\{\frac{1}{1} + \frac{h_1x_2}{1} + \frac{(1 - h_1)h_2x_3}{1} + \frac{(1 - h_2)h_3x_4}{1} + \cdots\right\}, \tag{6.9}$$

where $0 \leqq h_n < 1$ $(n = 1, 2, \cdots)$; *and if* $g_1 > 0$ *the series*

$$1 + \sum_{n=1}^{\infty} \frac{h_1h_2 \cdots h_n}{(1 - h_1)(1 - h_2) \cdots (1 - h_n)} \tag{6.10}$$

converges to the sum $\dfrac{1 - (1/S)}{g_1}$, *where S is the sum of the series* (6.2) (*possibly* ∞). *Thus, Theorem* 6.2 *is a consequence of Theorem* 6.1.

We shall suppose $g_1 > 0$ since the theorem is trivial if $g_1 = 0$. Put $x_2 = x_3 = \cdots = -1$ in (6.3), and denote the n-th numerator and denominator of the resulting continued fraction by P_n and Q_n, respectively. Then one may verify by mathematical induction that P_n and Q_n are polynomials in $g_1, g_2, \cdots, g_n$ given by the formulas $P_n = (1 - g_1)(1 - g_2) \cdots (1 - g_n)(S_n - 1)$, $Q_n = (1 - g_1)(1 - g_2) \cdots (1 - g_n)S_n$, where S_n is the sum of the first $n + 1$ terms of the series (6.2). If $s_n = (1 - g_n)g_{n+1}Q_{n-1}/Q_{n+1}$, then $s_n \geqq 0$ and the series $1 + \sum s_1s_2 \cdots s_n$ converges, its sum being equal to $\dfrac{1 - (1/S)}{g_1}$; $h_n = s_n/(1 + s_n)$ satisfies the inequality $0 \leqq h_n < 1$, and the series (6.10) converges to the sum $\dfrac{1 - (1/S)}{g_1}$. Moreover, one may verify at once that $h_1 = (1 - g_1)g_2$, $(1 - h_n)h_{n+1} = (1 - g_{n+1})g_{n+2}$ $(n = 1, 2, \cdots)$, and therefore (6.3) and (6.9) are equal, as was to be proved.

(iii) *If* $0 < g_n < 1$ $(n = 1, 2, \cdots)$ *and the series* (6.2) *diverges, then the continued fraction* (6.1) *converges for* $|x_n| \leqq 1$ $(n = 1, 2, \cdots)$, *except when* $x_1 = x_2 = \cdots = -1$.

Proof. If w is the value of (6.3), then (6.1) diverges if (and only if) $v = x_1 w = -1$. Since $|x_1| \leqq 1$, $|w| \leqq 1$, this implies that $|x_1| = 1$, $|w| = 1$. Now we may write: $w = g_1/[1 + (1 - g_1)v_1]$, where

$$v_1 = x_2\left\{\frac{g_2}{1} + \frac{(1 - g_2)g_3 x_3}{1} + \frac{(1 - g_3)g_4 x_4}{1} + \cdots\right\},$$

so that $|v_1| \leqq 1$. Consequently, we see that $|w - \frac{1}{2}| \leqq \frac{1}{2}$ and therefore $w = 1$, $x_1 = -1$, $v_1 = -1$. Repeating this argument starting with v_1 instead of with v, we find that $x_2 = -1$, and then, by mathematical induction, $x_3 = x_4 = \cdots = -1$.

Northwestern University.

SOME PROPERTIES OF SUMMABILITY

By J. D. Hill

Introduction. The purpose of this note is to present a few general results pertaining to some well-known and desirable properties of summability. By a method of summability, we shall ordinarily mean the familiar matrix method of assigning a limit to a sequence, although some of the broader remarks apply to any method of summability. The principal results are derived for the class of *reversible*[1] matrix methods, namely those for which the system of equations

$$\sum_{k=1}^{\infty} a_{mk}\, s_k = t_m \qquad (m = 1, 2, 3, \cdots)$$

has a unique solution $\{s_k\}$ corresponding to each convergent sequence $\{t_m\}$. The complex number system is employed throughout except where the contrary is specified.

1. **Translative methods.** We shall say that a method of summability is *translative to the right* (*to the left*) if the summability of the sequence $s_1, s_2, s_3, \cdots$ to the limit s always implies the summability of the sequence $s_2, s_3, s_4, \cdots$ ($s_0, s_1, s_2, \cdots$, for arbitrary s_0) to the limit s. A method that is both translative to the right and to the left will be called simply *translative*. We shall concern ourselves principally with methods of this latter type. It is immediately obvious that the summability of $s_1, s_2, s_3, \cdots$ to s by means of a translative method implies the summability of $s_{m+1}, s_{m+2}, s_{m+3}, \cdots$ to s, where m may be any positive or negative integer and s_k is to be interpreted as arbitrary if $k \leqq 0$. Furthermore, if the sequence of partial sums of the series $u_1 + u_2 + u_3 + \cdots$ is summable to s by means of a linear and regular translative method, then the sequences of partial sums of the series $u_2 + u_3 + u_4 + \cdots$ and $u_0 + u_1 + u_2 + \cdots$ (u_0 arbitrary) are summable to $s - u_1$ and $s + u_0$, respectively. Conversely, if the summability of the series $u_1 + u_2 + u_3 + \cdots$ to s by means of a linear regular method A always implies the simultaneous summability of the series $u_2 + u_3 + u_4 + \cdots$ and $u_0 + u_1 + u_2 + \cdots$ to $s - u_1$ and $s + u_0$, respectively, it follows that A must be translative.

The property described here as translativity has been mentioned by several writers[2] as a desirable adjunct to a method of summability, although the known

Received December 4, 1941.

[1] See S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932, p. 90. The term *reversible* has been used by I. Schur (*Über die Äquivalenz der Cesàroschen und Hölderschen Mittelwerte*, Mathematische Annalen, vol. 74(1913), pp. 447–458) in a sense different from that of Banach.

[2] The reader will find references easy to locate in L. L. Smail, *History and Synopsis of the Theory of Summable Infinite Processes*, University of Oregon Press, 1925.

results bear largely on special methods. Hardy,[3] for instance, has shown that Borel summability is translative to the left but not to the right. Knopp,[4] on the other hand, has established the fact that Euler summability is translative. Moreover, it is a trivial matter to verify directly that both Abel summability and regular Nörlund summability (of which Cesàro summability is a special case) are likewise translative methods. Since any method equivalent to a translative method is itself translative, the translativity of Hölder means follows from that of the Cesàro. Garabedian and Randels[5] have obtained a necessary and sufficient condition in order that regular Riesz means with positive weights shall be translative to the right. We shall show later (see Theorem 2) that this condition may be extended to complex weights, and in addition we obtain the condition for translativity to the left.

An approach to the general problem of translativity has been made by Hurwitz and Silverman,[6] and later by Carmichael.[7] The former obtain sufficient conditions in order that an analytically regular transformation shall be translative, and they also show that analytically regular transformations can be constructed which are translative neither to the left nor to the right. Carmichael goes a step further and states necessary and sufficient conditions in order that *normal*[8] and regular matrix summability shall be translative. We propose here to obtain necessary and sufficient conditions in order that the more general reversible matrix summability shall be translative. The conditions of Carmichael will appear as a special case, although in a more explicit form.

According to the definition stated above, the method of summability A defined by the matrix (a_{mk}) will be called translative if, for arbitrary s_0, the condition

$$(1.01) \qquad \lim_m \sum_{k=1}^{\infty} a_{mk} s_k \qquad \text{exists and equals } s$$

always implies the coexistence of

$$\lim_m \sum_{k=1}^{\infty} a_{mk} s_{k-1} \qquad \text{and} \qquad \lim_m \sum_{k=1}^{\infty} a_{mk} s_{k+1}$$

and their equality with s. As a consequence of this definition, we have at once the following lemma, whose proof we leave to the reader.[9]

[3] G. H. Hardy, *Researches in the theory of divergent series and divergent integrals*, Quarterly Journal of Mathematics, vol. 35(1904), pp. 22–66.

[4] K. Knopp, *Über das Eulersche Summierungsverfahren*, Mathematische Zeitschrift, vol. 15(1922), pp. 226–253.

[5] H. L. Garabedian and W. C. Randels, *Theorems on Riesz means*, Duke Mathematical Journal, vol. 4(1938), pp. 529–533. See Theorem 4, p. 532.

[6] W. A. Hurwitz and L. I. Silverman, *On the consistency and equivalence of certain definitions of summability*, Transactions of the American Mathematical Society, vol. 18(1917), pp. 1–20.

[7] R. D. Carmichael, *General aspects of the theory of summable series*, Bulletin of the American Mathematical Society, vol. 25(1918), pp. 97–131.

[8] In order that a *triangular* matrix (a_{mk}) be reversible, it is necessary and sufficient that a_{mm} be different from 0 for all m; such a matrix is called *normal*.

[9] Lemmas 1 and 2 were developed in collaboration with Prof. H. J. Hamilton.

LEMMA 1. *In order that A be translative, it is necessary and sufficient that*

$$\lim_m a_{mk} = 0 \qquad (k = 1, 2, 3, \cdots) \tag{1.02}$$

and that the condition (1.01) *always imply the coexistence of*[10]

$$\lim_m \sum_{k=1}^{\infty} a_{m,k-1}\, s_k \qquad \textit{and} \qquad \lim_m \sum_{k=1}^{\infty} a_{m,k+1}\, s_k$$

and their equality with s.

Let us denote by A_R and A_L the methods corresponding to the matrices $(a_{m,k-1})$ and $(a_{m,k+1})$, respectively. We may then express Lemma 1 in the following equivalent form.

LEMMA 2. *In order that A be translative, it is necessary and sufficient that condition* (1.02) *hold and that the methods A_R and A_L be not weaker than A and consistent with A.*

We may pause at this point to observe that if A is normal then the methods A_R and A_L will be not weaker than A and consistent with A if and only if the methods $A_R A^{-1}$ and $A_L A^{-1}$ are regular. The latter are essentially the conditions of Carmichael mentioned above.

In view of Lemma 2, the problem of characterizing translativity leads at once to the more general problem of determining necessary and sufficient conditions on a method B in order that it be not weaker than, and consistent with, a given method A. For the case in which A is assumed to be normal, a solution of the latter problem has been given by Mazur.[11] We proceed now to show that it is possible, by slight modifications of Mazur's proof, to obtain a similar result for the case in which A is assumed to be merely reversible. We recall that the method A corresponding to the matrix (a_{mk}) of complex numbers is said to be reversible if the system of equations

$$\sum_{k=1}^{\infty} a_{mk}\, s_k = t_m \qquad (m = 1, 2, 3, \cdots) \tag{1.03}$$

has a unique solution $\{s_k\}$ corresponding to each sequence $\{t_m\}$ in the space (c) of complex convergent sequences. In particular, if $\{t_m\}$ is allowed to become $Y \equiv \{\delta_m^m\}$ or $Y_n \equiv \{\delta_m^n\}$ for $n = 1, 2, 3, \cdots$, we denote the corresponding solutions $\{s_k\}$ by $\{\xi_k\}$ or $\{\xi_k^n\}$, respectively.

The main point in the proof of Mazur may be regarded as that of finding explicit expressions for the s_k in terms of the t_m and the *fundamental solutions* $\{\xi_k\}$ and $\{\xi_k^n\}$. In the event that A is normal, such expressions may be obtained very simply by means of Cramer's rule. In the present case we resort to the complex analogue of a theorem of Banach (see footnote 1, p. 47, Théorème 10), which, with regard to the application in view, may be stated as follows.

[10] We define a_{m0} as 0 $(m = 1, 2, 3, \cdots)$.

[11] S. Mazur, *Über lineare Limitierungsverfahren*, Mathematische Zeitschrift, vol. 28(1928), pp. 599–611; in particular, see Theorems III, IV, and V. We assume that the reader is familiar with this paper.

THEOREM A. *If the method A is reversible, the solution of the system* (1.03) *may be expressed in the form*

$$s_k(y) = C_k \lim_m t_m + \sum_{m=1}^{\infty} C_{mk} t_m \qquad (k = 1, 2, 3, \cdots) \tag{1.04}$$

in which the coefficients C_k and C_{mk} are independent of $y \equiv \{t_m\}$ and satisfy the condition $\sum_{m=1}^{\infty} | C_{mk} | < \infty$ $(k = 1, 2, 3, \cdots)$.

The proof of this theorem follows exactly the lines of the original as far as the final remark, which states that the $s_k(y)$ are linear functionals in the space of the sequences $y = \{t_m\}$. In the present case, it is easily seen that the $s_k(y)$ are additive and continuous operations on the space (c) to the space of complex numbers that satisfy the condition $s_k(zy) = z \cdot s_k(y)$ for every complex number z. The derivation of the general form of such an operation is formally identical with the derivation of the general linear functional in the space of real convergent sequences (see Banach, footnote 1, p. 65, §3). It is clear then that $s_k(y)$ may be written in the form (1.04), and since we have $\| s_k \| = | C_k | + \sum_{m=1}^{\infty} | C_{mk} |$, the final remark in the theorem is verified.

It remains to express the coefficients C_k and C_{mk} in terms of the fundamental solutions $\{\xi_k\}$ and $\{\xi_k^n\}$ defined above. We have $\xi_k^n \equiv s_k(Y_n) = C_{nk}$, and consequently $\xi_k \equiv s_k(Y) = C_k + \sum_{m=1}^{\infty} \xi_k^m$. If we now define ξ_k^0 and t_0 by means of the relations $\xi_k^0 \equiv C_k = \xi_k - \sum_{m=1}^{\infty} \xi_k^m$ and $t_0 \equiv \lim_m t_m$, then (1.04) may be written in the form

$$s_k(y) = \sum_{m=0}^{\infty} \xi_k^m t_m \qquad (k = 1, 2, 3, \cdots). \tag{1.05}$$

On the basis of equations (1.05), it is now possible, with no essential modifications, to follow the lines of Mazur's proof and thus arrive at the following theorem.

THEOREM B. *Let A with matrix (a_{mk}) be a given reversible method, and let B with matrix (b_{mk}) be any method whatsoever. Then in order that B be not weaker than A and consistent with A the following conditions are necessary and sufficient.*

$$B\text{-}\lim \{\xi_k\} \textit{ exists and } = 1; \tag{1.06}$$

$$B\text{-}\lim \{\xi_k^n\} \textit{ exists and } = 0 \qquad (n = 1, 2, 3, \cdots); \tag{1.07}$$

$$\sup_r \sum_{n=1}^{\infty} \left| \sum_{k=1}^{r} b_{mk} \xi_k^n \right| < \infty \qquad (m = 1, 2, 3, \cdots); \tag{1.08}$$

$$\sup_m \sum_{n=1}^{\infty} \left| \sum_{k=1}^{\infty} b_{mk} \xi_k^n \right| < \infty. \tag{1.09}$$

Returning now to Lemma 2 and taking into account the result of Theorem B, we may state the following characterization of reversible translative methods.

THEOREM 1. *In order that a reversible method A be translative, it is necessary and sufficient that condition* (1.02) *hold and that the methods A_R and A_L satisfy the conditions* (1.06)–(1.09).

As an application of Theorem 1, we take A as the normal and regular method of Riesz means. The latter in its most general form is defined by a triangular matrix whose elements are given by $a_{mk} \equiv p_k/P_m$ $(k = 1, 2, 3, \cdots, m;$ $m = 1, 2, 3, \cdots)$, where the complex numbers p_k and $P_m \equiv p_1 + p_2 + \cdots + p_m$ $(k, m = 1, 2, 3, \cdots)$ are all different from 0 and satisfy the Silverman-Toeplitz regularity conditions. We notice that condition (1.02) is implied by the regularity.

One may readily verify in this case that the fundamental solutions are given by the following equations.

$$\xi_k = 1 \qquad (k = 1, 2, 3, \cdots);$$

$$\xi_n^n = \frac{P_n}{p_n}, \qquad \xi_{n+1}^n = \frac{-P_n}{p_{n+1}}; \qquad \xi_k^n = 0 \text{ otherwise} \qquad (k, n = 1, 2, 3, \cdots). \tag{1.10}$$

Furthermore, since the regularity of A implies that of A_R and A_L, it is clear from (1.10) that both A_R and A_L satisfy the conditions (1.06) and (1.07).

We consider next the conditions (1.08) and (1.09) as applied to A_R and A_L. By employing (1.10) to evaluate the summations involved, we find without difficulty the following expressions, wherein p_0 is understood to be 0.

$$G_{mr}^R \equiv \sum_{n=1}^{\infty} \left| \sum_{k=1}^{r} a_{m,k-1} \xi_k^n \right| = \frac{1}{|P_m|} \left\{ \sum_{n=1}^{r-1} | p_{n-1} \xi_n^n + p_n \xi_{n+1}^n | + | p_{r-1} \xi_r^r | \right\}; \tag{1.11}$$

$$G_{mr}^L \equiv \sum_{n=1}^{\infty} \left| \sum_{k=1}^{r} a_{m,k+1} \xi_k^n \right| = \frac{1}{|P_m|} \left\{ \sum_{n=1}^{r-1} | p_{n+1} \xi_n^n + p_{n+2} \xi_{n+1}^n | + | p_{r+1} \xi_r^r | \right\}; \tag{1.12}$$

$$H_m^R \equiv \sum_{n=1}^{\infty} \left| \sum_{k=1}^{\infty} a_{m,k-1} \xi_k^n \right| = \frac{1}{|P_m|} \sum_{n=1}^{m} | P_n | \left| \frac{p_{n-1}}{p_n} - \frac{p_n}{p_{n+1}} \right| + \left| \frac{p_m P_{m+1}}{p_{m+1} P_m} \right|; \tag{1.13}$$

$$H_m^L \equiv \sum_{n=1}^{\infty} \left| \sum_{k=1}^{\infty} a_{m,k+1} \xi_k^n \right| = \frac{1}{|P_m|} \sum_{n=1}^{m-2} | P_n | \left| \frac{p_{n+1}}{p_n} - \frac{p_{n+2}}{p_{n+1}} \right| + \left| \frac{p_m P_{m-1}}{p_{m-1} P_m} \right|. \tag{1.14}$$

Since a_{mk} in this case is 0 for $k > m$, we see from (1.11) and (1.12) that for each fixed $m = 1, 2, 3, \cdots$

$$G_{mr}^R = G_{m,m+1}^R \qquad (r = m + 2, m + 3, m + 4, \cdots),$$

$$G_{mr}^L = G_{m,m-1}^L \qquad (r = m, m + 1, m + 2, \cdots),$$

from which it follows that A_R and A_L satisfy the condition (1.08).

With the remaining condition, the state of affairs is complicated by the fact that regularity alone is not sufficient to ensure the satisfaction of (1.09). Consider the following examples.

Example 1. Let $p_{3n-2} = 1/n$, $p_{3n-1} = 1/n!$, $p_{3n} = 1/(n+1)!$ for $n = 1, 2, 3, \cdots$. The corresponding method (R, p_k) is normal and regular, but from (1.13) we have

$$H^R_{3n-1} > \frac{p_{3n-1} P_{3n}}{p_{3n} P_{3n-1}} = (n+1) \frac{P_{3n}}{P_{3n-1}} > n + 1.$$

Example 2. Let $p_{3n-2} = (n+1)!$, $p_{3n-1} = n!$, $p_{3n} = (n+1)!$ for $n = 1, 2, 3, \cdots$. The method (R, p_k) so defined is normal and regular, but (1.14) gives

$$H^L_{3n} > \frac{p_{3n} P_{3n-1}}{p_{3n-1} P_{3n}} = (n+1)\cdot\left(1 - \frac{p_{3n}}{P_{3n}}\right) > \frac{n+1}{2},$$

since $P_{3n} > 2(n+1)!$.

In view of the preceding examples, we see that Theorem 1 in the present instance reduces to the following form.

THEOREM 2. *In order that a normal and regular method of Riesz means be translative it is necessary and sufficient that the sequences* $\{H^R_m\}$ *and* $\{H^L_m\}$ *be bounded.*

This theorem improves the result of Garabedian and Randels to the extent previously mentioned. These writers (see footnote 5, p. 532) have also observed that for regular Riesz means with positive weights the monotonicity in either sense of the sequence $\{p_n/p_{n+1}\}$ is sufficient for translativity to the right. It is to be noticed that in the same circumstances this condition is likewise sufficient for translativity to the left.

The method (R, p_k) defined above in Example 1 may be used to illustrate the curious behavior of non-translative methods. For, let $s_1, s_2, s_3, \cdots$ denote the sequence $0, 0!, 0, 0, 1!, 0, \cdots, 0, (k-1)!, 0, \cdots$. One easily shows that this sequence is summable-(R, p_k) to 1, whereas the sequence $s_0, s_1, s_2, \cdots$ is summable-(R, p_k) to 0, and the sequence $s_2, s_3, s_4, \cdots$ is such that its (R, p_k)-transform diverges to $+\infty$. On the other hand, let $t_1, t_2, t_3, \cdots$ denote the sequence $1, 0, 0, 1, 0, 0, \cdots, 1, 0, 0, \cdots$. This sequence is summable-(R, p_k) to 1, while each of the sequences $t_0, t_1, t_2, \cdots$ and $t_2, t_3, t_4, \cdots$ is summable-(R, p_k) to 0.

Finally, we may point out that the notion of translativity enters naturally if we attempt to generalize the fact that the terms of a convergent series form a null sequence. If we ask to what extent this property is preserved when convergence is replaced by summability, we find an answer in the following theorem.

THEOREM 3. *In order that the summability of the series* $\sum_1^\infty u_k$ *by a linear regular method A shall always imply the A-summability of the sequence* $\{u_k\}$ *to* 0, *it is necessary and sufficient that A be translative to the left.*

The proof follows at once from the identity $(u_1, u_2, u_3, \cdots) \equiv (s_1, s_2, s_3, \cdots) - (s_0, s_1, s_2, \cdots) + (s_0, 0, 0, \cdots)$, where $s_k \equiv u_1 + u_2 + \cdots + u_k$ $(k \geqq 1)$ and s_0 is arbitrary.

2. **Methods stronger than convergence.** In order that a regular method of summability A should constitute a non-redundant generalization of ordinary convergence, it is necessary that at least one divergent sequence should be summable-A. If this condition is satisfied, we shall say that A is *stronger than convergence*. When faced with the problem of establishing this property for a given method A, one naturally attempts to construct a divergent sequence that is summable-A. The success of this process in practice is well known. On the other hand, it may be of interest to observe that it is possible to state a characterization of the reversible regular methods for which the preceding property holds. For, let us consider the reversible regular transformation

$$\sum_{k=1}^{\infty} a_{mk} s_k = t_m \qquad (m = 1, 2, 3, \cdots) \tag{2.1}$$

and its inverse transformation

$$\sum_{m=0}^{\infty} \xi_k^m t_m = s_k \qquad (k = 1, 2, 3, \cdots) \tag{2.2}$$

as found in §1. It is evident that A will be stronger than convergence if and only if the transformation (2.2), regarded as a transformation of the space (c), *fails* to be convergence-preserving. Now a transformation of the type (2.2) differs from one of the type (2.1) in that the former involves t_0, the limit of the sequence $\{t_m\}$. However, it is easy to see that the familiar necessary and sufficient conditions of Schur for preservation of convergence in (2.1) require but a minor modification in order to hold for (2.2). We find therefore that (2.2) will be convergence-preserving if and only if the following conditions are satisfied.

$$\sup_k \sum_{m=0}^{\infty} | \xi_k^m | < \infty ; \tag{2.3}$$

$$\lim_k \sum_{m=0}^{\infty} \xi_k^m \text{ exists}; \tag{2.4}$$

$$\lim_k \xi_k^m \text{ exists} \qquad (m = 1, 2, 3, \cdots). \tag{2.5}$$

From these remarks, we conclude the truth of the following theorem.

THEOREM 4. *In order that a reversible regular method be stronger than convergence, it is necessary and sufficient that at least one of the conditions* (2.3), (2.4), (2.5) *be violated.*

In applying this theorem to normal regular Riesz means (see §1) we find that (2.4) and (2.5) are always satisfied. Condition (2.3) must therefore fail to hold, and this reduces to the condition

$$\overline{\lim_k} \frac{| P_{k-1} | + | P_k |}{| p_k |} = +\infty . \tag{2.6}$$

If $p_k > 0$ for all k, condition (2.6) becomes simply $\overline{\lim}_k P_k/p_k = +\infty$.

3. **Methods of type** M. The method of summability A defined by the matrix (a_{mk}) is said to be of[12] *type* M if the conditions (see Banach, footnote 1)

$$\sum_{m=1}^{\infty} |\alpha_m| < \infty, \qquad \sum_{m=1}^{\infty} \alpha_m a_{mk} = 0 \qquad (k = 1, 2, 3, \cdots) \tag{3.1}$$

always imply

$$\alpha_m = 0 \qquad (m = 1, 2, 3, \cdots). \tag{3.2}$$

The significance of this property is shown by the following theorems.

THEOREM OF MAZUR.[13] *In order that a normal regular method A be consistent with every regular method not weaker than A, it is necessary and sufficient that A be of type M.*

THEOREM OF BANACH.[14] *In order that a reversible regular method A be consistent with every regular method not weaker than A, it is sufficient that A be of type M.*

Banach (see footnote 1, p. 236, lines 33–36) remarks without proof that the type M condition in the latter theorem is necessary as well as sufficient. Since no proof of this fact seems to have appeared in the literature the following one may be of interest.

THEOREM 5. *In order that a reversible regular method A be consistent with every regular method not weaker than A, it is necessary and sufficient that A be of type M.*

Proof of necessity. We assume that A is a reversible regular method consistent with every regular method not weaker than A and that $\{\alpha_m\}$ is an arbitrary sequence satisfying the conditions (3.1). We have to show that conditions (3.2) necessarily follow. With this in view, let $b_{nk} \equiv \sum_{m=1}^{n} \alpha_m a_{mk}$ $(k, n = 1, 2, 3, \cdots)$ and let B denote the method defined by the matrix (b_{nk}). If $\{s_k\}$ is an arbitrary A-summable sequence, we set $\sum_{k=1}^{\infty} a_{mk} s_k = t_m$ $(m = 1, 2, 3, \cdots)$, so that the sequence $\{t_m\}$ is convergent and therefore bounded. It follows easily then that

$$\lim_n \sum_{k=1}^{\infty} b_{nk} s_k = \lim_n \sum_{m=1}^{n} \alpha_m t_m = \sum_{m=1}^{\infty} \alpha_m t_m. \tag{3.3}$$

[12] See J. D. Hill, *On perfect methods of summability*, Duke Mathematical Journal, vol. 3(1937), pp. 702–714. This paper is devoted to methods of type M that are regular and reversible; such methods are called *perfect*.

[13] S. Mazur, *Eine Anwendung der Theorie der Operationen bei der Untersuchung der Toeplitzschen Limitierungsverfahren*, Studia Mathematica, vol. 2(1930), p. 48, Satz 7.

[14] See footnote 1, p. 95, Théorème 12. The proof is given for the real domain but it may be easily extended to the complex.

This relation shows that every A-summable sequence is summable-B to the value $\sum_{m=1}^{\infty} \alpha_m t_m$. Furthermore, if $\{s_k\}$ is convergent, we make use of (3.1) and the fact that the double series $\sum_{k,m=1}^{\infty} \alpha_m a_{mk} s_k$ converges absolutely to show that

$$\sum_{m=1}^{\infty} \alpha_m t_m = \sum_{m=1}^{\infty} \sum_{k=1}^{\infty} \alpha_m a_{mk} s_k = \sum_{k=1}^{\infty} \left(\sum_{m=1}^{\infty} \alpha_m a_{mk} \right) s_k = 0.$$

Consequently, every convergent sequence is summable-B to 0.

It is now apparent that the method C defined by the matrix $(a_{mk} + b_{mk})$ is regular and not weaker than A. Moreover, since A is reversible, we infer the existence of a sequence $\{\sigma_k\}$ satisfying the equations

$$\sum_{k=1}^{\infty} a_{mk} \sigma_k = \bar{\alpha}_m \qquad (m = 1, 2, 3, \cdots),$$

where $\bar{\alpha}_m$ is the complex conjugate of α_m. Since $\lim_m \alpha_m = 0$, the sequence $\{\sigma_k\}$ is summable-A to 0; and in view of (3.3), it is summable-C to the value $\sum_{m=1}^{\infty} | \alpha_m |^2$. If A and C are consistent, the latter expression must vanish, and this completes the proof.

MICHIGAN STATE COLLEGE.

A GENERAL EQUATION FOR RELAXATION OSCILLATIONS

By Norman Levinson and Oliver K. Smith

1. **Introduction.** The importance of relaxation oscillations in physical and engineering problems was shown by Van der Pol,[1] who also treated by graphical methods a particular equation describing these oscillations. The origin of relaxation oscillations can be described *qualitatively* by considering the following two equations with constant coefficients:

$$\ddot{x} + 2a\dot{x} + bx = 0, \tag{1.1}$$

$$\ddot{x} - 2a\dot{x} + bx = 0, \tag{1.2}$$

where $a > 0$, $b > a^2$, and the differentiations are with respect to t. If we denote $b - a^2$ by ω^2 then the solution of (1.1) is $Ae^{-at} \sin (\omega t + \alpha)$ and the solution of (1.2) is $Ae^{at} \sin (\omega t + \alpha)$, where A and α are arbitrary constants. In an electrical circuit, for example, described by (1.1), the term $2a\dot{x}$ arises from a withdrawal of energy from the system. Since there is no energy being put into the system, this withdrawal is uncompensated and results in a gradual dissipation of the initial energy of the system. Thus the solution of (1.1) tends to zero. In (1.2), on the other hand, the only term affecting the energy is $-2a\dot{x}$ which arises from adding energy to the system. Thus the solution of (1.2) describes oscillations of every increasing amplitude.

The equation (1.1) may be said to describe a system with positive damping and (1.2), a system with negative damping. Positive damping decreases the energy and therefore the amplitude of an oscillation while negative damping increases it. Relaxation oscillations arise when both positive and negative damping occur in a system. More precisely what occurs is that for small displacements, that is, when x is small, the system has negative damping which causes oscillations of increasing amplitude; on the other hand, for large displacements the system has positive damping which tends to decrease the amplitude of oscillation. Clearly, then, the steady state amplitude of oscillation of the system cannot be too small, for then damping would always be negative and the oscillation would increase in amplitude, that is, would not be steady-state. In the same way a very large amplitude is not possible. Thus qualitatively one would expect a steady-state oscillation of such amplitude that during each period the energy lost when the displacement was large (and damping positive) would be exactly compensated by the energy gained when the displacement was small (and damping negative).

All this is what one would expect qualitatively. Actually, under a wide range of conditions precisely this does happen. On the other hand, there are also

Received December 10, 1941.

[1] B. Van der Pol, *Relaxation-Oscillations*, Philosophical Magazine, seventh series, vol. 2(1926), p. 978.

equations satisfying our qualitative description which do not have any steady-state oscillating solution at all. In any case a mere qualitative discussion is entirely inconclusive. Von Kármán[2] has, in fact, stated that the question of existence of solutions for non-linear problems is of considerable practical interest since the existence of such solutions is by no means obvious.

The equation considered in detail by Van der Pol was

$$\ddot{x} - \mu(1 - x^2)\dot{x} + x = 0.$$

With $\mu > 0$, it is clear that for $|x| < 1$ the damping is negative and for $|x| > 1$, positive. From graphical considerations Van der Pol obtained the solutions for various values of μ and in each case the result was a rapid approach to a steady-state oscillation.

A more general description of relaxation oscillations is given by the equation

$$\ddot{x} + f(x)\dot{x} + x = 0, \tag{1.3}$$

where $f(x)$ is negative for small values of $|x|$ and positive for large values of $|x|$. A case of particular importance is where $f(x)$ is an even function of x. For this case the first satisfactory treatment of (1.3) is due to Liénard.[3] Liénard introduces

$$F(x) = \int_0^x f(x)\,dx.$$

Since $f(x)$ is even, $F(x)$ is an odd function. Liénard shows that if there exists some value of x, x_0, such that, for $0 < x < x_0$, $F(x)$ is negative while, for $x > x_0$, $F(x)$ is positive and steadily increasing with $F(x) \to \infty$ as $x \to \infty$, then all solutions of (1.3) tend toward a steady oscillatory solution which is unique except for its phase. (This latter is to be expected since (1.3) is unchanged by a translation of the time scale, and thus translations of a solution are also solutions.) Liénard's work represents the furthest progress in this problem to date.

Here we propose to consider the generalized equation for relaxation oscillations

$$\ddot{x} + f(x, \dot{x})\dot{x} + g(x) = 0. \tag{1.4}$$

$g(x)$ is positive when $x > 0$ and negative when $x < 0$. $f(x, \dot{x})$ is the damping coefficient which for large $|x|$ is positive and for small $|\dot{x}|$ and $|x|$ is negative. With little more than these requirements we shall show that (1.4) possesses periodic solutions and is therefore a generalized equation for relaxation oscillations. The question of the existence of a unique periodic solution is more involved and can be settled only with further restrictions. This problem we

[2] T. Von Kármán, *The Engineer Grapples with Nonlinear Problems*, Bulletin of the American Mathematical Society, vol. 46(1940), p. 617.

[3] A. Liénard, *Etude des oscillations entretenues*, Revue Gén. de l'Electricité, t. XXIII(1928), pp. 901–946.

shall also consider. As a particular case of this consideration the existence of a unique solution for

$$\ddot{x} + f(x)\dot{x} + g(x) = 0 \tag{1.5}$$

will be demonstrated when $f(x)$ is not necessarily even nor $g(x)$ necessarily odd.

Using the method of Liénard we shall also demonstrate the existence of a unique steady-state oscillating solution for (1.5) under precisely the same conditions on $f(x)$ as was given above in the statement of Liénard's result for (1.3).

2. **On the existence of periodic solutions.** Here we shall show that the equation

$$\ddot{x} + f(x, \dot{x})\dot{x} + g(x) = 0 \tag{2.0}$$

possesses periodic solutions and thereby justifies our calling it a generalized equation for relaxation oscillations. The equation (2.0) can be written as a pair of first order equations.

$$\frac{dx}{dt} = v, \qquad \frac{dv}{dt} = -f(x, v)v - g(x). \tag{2.1}$$

We shall assume throughout that the derivative of $g(x)$ and the first order partial derivatives with respect to x and v of $f(x, v)$ exist and are continuous. Since dt can be eliminated from the pair of equations (2.1) resulting in a first order equation in x and v, it follows that to solutions of (2.0) correspond curves in the (x, v)-plane which are solutions of the first order equation in x and v. In particular, to a periodic solution of (2.0) there must correspond a closed curve in the (x, v)-plane since x and v return to their initial values in the course of the period of such a solution. Conversely, any solution of (2.1), which, when considered as a curve in the (x, v)-plane with parameter t, is a closed curve traversed with a finite change in t, corresponds to a periodic solution of (2.0). Thus the problem of finding periodic solutions of (2.0) is reduced to the problem of showing that there are solutions (2.1) forming closed curves in the (x, v)-plane, which are traversed with a finite change in the parameter t.

We now state our theorem in precise terms.

THEOREM I. *Let $xg(x) > 0$ for $|x| > 0$. Moreover let*

$$\int_0^{\pm\infty} g(x)\, dx = \infty. \tag{2.2}$$

Let $f(0, 0) < 0$ and let there exist some $x_0 > 0$ such that $f(x, v) \geqq 0$ for $|x| \geqq x_0$. Further, let there exist an M such that for $|x| \leqq x_0$

$$f(x, v) \geqq -M. \tag{2.3}$$

Finally, let there exist some $x_1 > x_0$ such that[4]

[4] In (2.4) it would of course be equally good to have the integration from $(-x_1, -x_0)$ and then to take v as negative.

$$\int_{x_0}^{x_1} f(x, v)\, dx \geqq 10Mx_0, \tag{2.4}$$

where $v > 0$ is an arbitrary decreasing positive function of x in the integration in (2.4). *Under these conditions* (2.0) *has at least one periodic solution.*

In the proof of this theorem we make considerable use of the function

$$G(x) = \int_0^x g(x)\, dx.$$

Since $xg(x) > 0$, it follows that $G(x) > 0$ for $x \neq 0$ and that $G(x)$ decreases monotonically as x increases from $-\infty$ to 0 and increases monotonically as x increases from 0 to ∞. From (2.2) it follows further that $\lim G(x)$ ($|x| \to \infty$) is ∞. This last result can be replaced by a weaker requirement which states merely that $G(x)$ gets large for large x. Thus instead of (2.2), we first require that, for some $x_1 > x_0$,

$$G(x_1) - G(x_0) \geqq \max\left(400M^2x_0^2, \frac{G^2(x_0)}{M^2x_0^2}\right). \tag{2.5}$$

Since $G(x)$ is an increasing function for positive x, it follows that if (2.5) is satisfied it continues to be satisfied if we increase x_1. Thus x_1 in (2.5) can be chosen so as to be at least as big as x_1 in (2.4). Having so chosen x_1, the only requirement beyond (2.5) needed to replace (2.2) is that

$$\lim_{|x|\to\infty} G(x) \geqq 2G(x_1). \tag{2.6}$$

Since (2.4) continues to be satisfied if x_1 is increased by virtue of the fact that $f(x, v) \geqq 0$ for $x > x_0$, it follows finally that we can, if necessary, increase x_1 in (2.4) so that x_1 in (2.4), (2.5) and (2.6) are all equal.

Before proceeding to the proof of Theorem I we shall consider the general pair of equations

$$\frac{dx}{dt} = X(x, y), \qquad \frac{dy}{dt} = Y(x, y), \tag{2.7}$$

where X and Y possess continuous first order partial derivatives in x and y. The equations (2.7) include as a particular case the equations (2.1) and therefore are of interest to us here. The equations (2.7) have been treated in considerable detail and we shall merely state some of their properties. Those points (x, y) at which both X and Y vanish are known as singular points. Through every point in the (x, y)-plane with the possible exception of singular points there passes one and only one solution of (2.7), this solution being given in terms of the parameter t. Such a solution forms a part of an integral curve. An integral curve cannot cross itself or another integral curve except at a singular point. Moreover, the change in t in going between two points on an integral curve is always finite unless either of these points or some intermediate point on the curve is a singular point. Conversely, if an integral curve runs into a singular point the change in t as the point is approached along the curve tends to infinity.

Now let us consider an integral curve which for $t \to \infty$ remains in a finite region R in the (x, y)-plane and let R be free of singular points. Then, as $t \to \infty$, the length of the curve must tend to infinity for if the curve is finite it terminates in a point as $t \to \infty$, which is impossible in a region free of singular points.

We have then an infinite curve in a finite region which never intersects itself. Sketching a few such curves will make very plausible indeed the following theorem.

THEOREM A.[5] *If an integral curve of* (2.7) *lies in a finite region R for $t \to \infty$ and if there are no singular points in R, then the integral curve is either a closed curve or else it approaches nearer and nearer to a closed integral curve.*

Thus returning to (2.1) we have a general method[6] for demonstrating the existence of closed integral curves in the (x, v)-plane or, in other words, periodic solutions of (2.0). We now proceed to demonstrate the existence of a region in the (x, v)-plane satisfying the requirements of Theorem A, thereby proving Theorem I with the less restrictive (2.5) and (2.6) in place of (2.2).

Proof of Theorem I. In the equations (2.1) the only singular point in the (x, v)-plane is $(0, 0)$. For, clearly, $dx/dt = v$ vanishes only if $v = 0$. Once $v = 0$, the other equation becomes $dv/dt = g(x)$. But $g(x) = 0$ only at $x = 0$. Thus $(0, 0)$ is the only singular point.

We now introduce

$$\lambda(x, v) = \tfrac{1}{2}v^2 + G(x). \tag{2.8}$$

Since $G'(x) = g(x)$, it follows that the curves $\lambda(x, v) = c$ must have negative slope when x and v are both positive or both negative, and have positive slope otherwise. Clearly

$$\frac{d\lambda}{dt} = v\frac{dv}{dt} + g(x)\frac{dx}{dt} = v\left(\frac{dv}{dt} + g(x)\right).$$

Or, using the equation in (2.1) for dv/dt, this becomes

$$\frac{d\lambda}{dt} = -v^2 f(x, v). \tag{2.9}$$

Thus if $f(x, v) > 0$, then as t increases the integral curves of (2.1) in the (x, v)-plane cut across the curves $\lambda(x, v) = c$ so that λ decreases while if $f(x, v) < 0$ the integral curves cut the $\lambda(x, v) = c$ curves so that λ increases. In par-

[5] All the results stated above as well as Theorem A are proved in Ivar Bendixson, *Sur les courbes définies par les équations differentielles*, Acta Mathematica, vol. 24(1901), pp. 1–88.

[6] This method has been used by one of the authors, O. K. Smith, in his thesis at M. I. T. (May, 1941) to prove Theorem I in a more restrictive form and with $g(x) = x$. The method has also been used by V. S. Ivanov in a paper which is reviewed in Math. Reviews, vol. 2 (1941), p. 287. According to the review, however, Ivanov merely applies the method to the equation dealt with by Liénard.

ticular, since $f(0, 0) < 0$, around the origin the integral curves cut outward across $\lambda(x, v) = c$. Thus the integral curves move outward from the origin (which is the only singular point) as t increases. (See AB in Figure 1.) On the other hand, since for $|x| \geqq x_0$, $f(x, v) \geqq 0$, it follows that integral curves for which $|x| > x_0$ cut inward across the curves $\lambda(x, v) = c$ as in Figure 1. (The direction of the curves sketched, CD moving to the right and EF to the left, comes from the equation $dx/dt = v$. When $v > 0$, this equation states that x increases as t increases whereas, when $v < 0$, x decreases as t increases.)

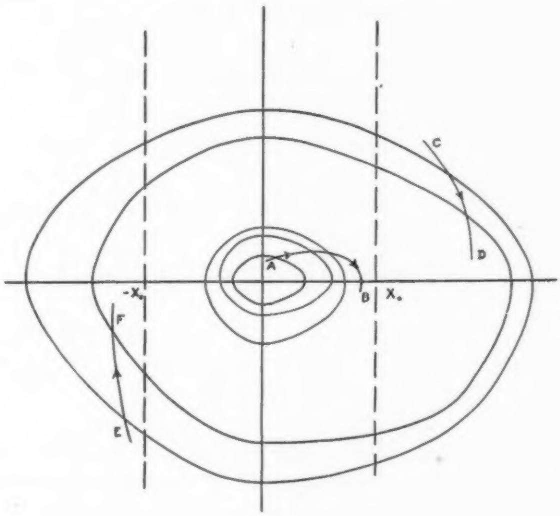

FIG. 1

We now introduce

$$v_0 = 2(G(x_1) - G(x_0))^{\frac{1}{2}} \tag{2.10}$$

and consider the curve

$$\lambda(x, v) = \tfrac{1}{2}v_0^2 + G(x_0) = \lambda_0 . \tag{2.11}$$

Clearly,

$$\lambda_0 = 2G(x_1) - G(x_0) < 2G(x_1). \tag{2.12}$$

The curve $\lambda(x, v) = \lambda_0$ will be closed if it intersects the positive and negative x-axis. And this latter depends on the possibility of solving $\lambda(x, 0) = \lambda_0$, or, since $\lambda(x, 0) = G(x)$, on solving $G(x) = \lambda_0$ for positive and negative values of x. But $\lambda_0 < 2G(x_1)$ and, by (2.6), $G(x)$ for sufficiently large $|x|$ exceeds $2G(x_1)$. Since $G(x)$ increases with $|x|$ and $G(0) = 0$, it follows that $\lambda(x, 0) = \lambda_0$ has a solution for positive and for negative x. Thus the curve $\lambda(x, v) = \lambda_0$ is closed.

This curve is *abcd* of Figure 2. We next consider the solution of (2.1) which starts at the point (x_0, v_0), point A in the figure. We have seen that for $|x| \geqq x_0$, $d\lambda/dt \leqq 0$ along an integral curve and thus the integral curve we are considering cuts inward across the curves $\lambda(x, v) = c$ so long as $x > x_0$. Let us suppose the integral curve through A intersects the line $x = x_1$ for the first time at B. We denote the value of $\lambda(x, v)$ at B by λ_1. From (2.9) and $dx = v\,dt$ it follows that

$$\frac{d\lambda}{dx} = -f(x, v)v. \tag{2.13}$$

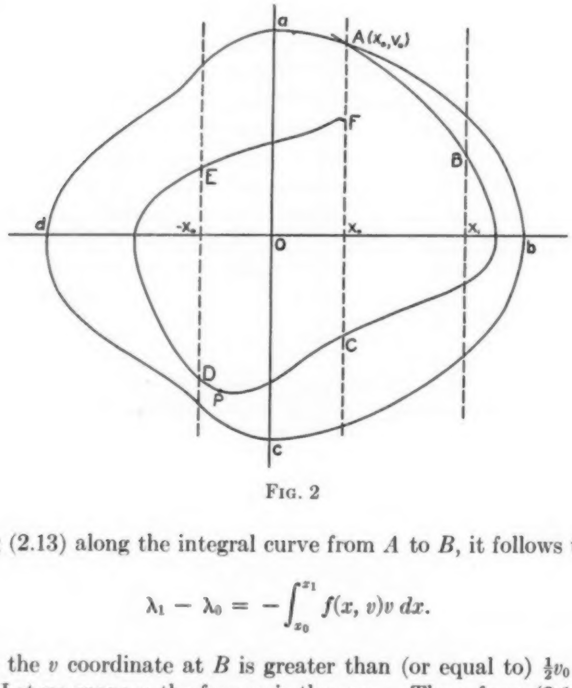

FIG. 2

Integrating (2.13) along the integral curve from A to B, it follows that

$$\lambda_1 - \lambda_0 = -\int_{x_0}^{x_1} f(x, v)v\,dx. \tag{2.14}$$

Now either the v coordinate at B is greater than (or equal to) $\frac{1}{2}v_0$ or it is less than $\frac{1}{2}v_0$. Let us suppose the former is the case. Then, from (2.14),

$$\lambda_1 - \lambda_0 \leqq -\tfrac{1}{2}v_0 \int_{x_0}^{x_1} f(x, v)\,dx$$

and, using (2.4),

$$\lambda_1 - \lambda_0 \leqq -\tfrac{1}{2}v_0(10Mx_0).$$

Or

$$\lambda_1 \leqq \lambda_0 - 5Mx_0v_0. \tag{2.15}$$

On the other hand, if v at B is less than $\frac{1}{2}v_0$, then

$$\lambda_1 \leqq \frac{1}{2}\left(\frac{v_0}{2}\right)^2 + G(x_1).$$

Or, if we use the value of λ_0 in (2.12),

$$\begin{aligned}\lambda_1 &\leqq \tfrac{1}{2}[G(x_1) - G(x_0)] + G(x_1)\\ &= \tfrac{3}{2}G(x_1) - \tfrac{1}{2}G(x_0)\\ &= \tfrac{3}{4}\lambda_0 + \tfrac{3}{4}G(x_0) - \tfrac{1}{2}G(x_0)\\ &= \tfrac{3}{4}\lambda_0 + \tfrac{1}{4}G(x_0).\end{aligned}$$

Thus

$$\begin{aligned}\lambda_1 &\leqq \lambda_0 - \tfrac{1}{4}[\lambda_0 - G(x_0)]\\ &= \lambda_0 - \frac{1}{4}\frac{v_0^2}{2} \qquad (2.16)\\ &= \lambda_0 - \frac{v_0}{4}\,(G(x_1) - G(x_0))^{\frac{1}{2}}.\end{aligned}$$

Using (2.5), this becomes

$$\lambda_1 < \lambda_0 - \frac{v_0}{4}\,(20Mx_0) = \lambda_0 - 5Mx_0v_0\,.$$

Thus, in any case, (2.15) holds if the integral curve intersects the line $x = x_1$. Since λ continues to decrease along the integral curve for $x > x_0$, it follows that if λ at C (in Figure 2) on the integral curve be denoted by λ_2 then $\lambda_2 \leqq \lambda_1$ and, therefore,

$$\lambda_2 \leqq \lambda_0 - 5M_0x_0v_0\,. \qquad (2.17)$$

In case the integral curve does not intersect the line $x = x_1$, it follows that it must cut the positive x-axis between x_0 and x_1 at, let us say, x_2. At this point $(x_2, 0)$, $\lambda = G(x_2)$. But since $G(x)$ grows with $|x|$, $G(x_2) < G(x_1)$ and, therefore, at $(x_2, 0)$, $\lambda < G(x_1)$. Since λ decreases as the integral curve proceeds to C, it follows that λ at (C, λ_2) is less than λ at $(x_2, 0)$. Thus, in this case,

$$\begin{aligned}\lambda_2 < G(x_1) &= \tfrac{1}{2}\lambda_0 + \tfrac{1}{2}G(x_0)\\ &= \lambda_0 - \tfrac{1}{2}[\lambda_0 - G(x_0)].\end{aligned}$$

Proceeding as in (2.16),

$$\lambda_2 \leqq \lambda_0 - 10Mx_0v_0 < \lambda_0 - 5Mx_0v_0\,.$$

Thus in any case (2.17) holds.

We now assume that the integral curve cuts the line $x = -x_0$ at D. Denoting λ at D by λ_3 and integrating (2.13) from C to D,

$$\lambda_3 - \lambda_2 = -\int_{x_0}^{-x_0} vf(x, v)\, dx$$

or

$$\lambda_3 - \lambda_2 = -\int_{x_0}^{-x_0} |v| f(x, v)\, dx$$

$$\leqq M \int_{-x_0}^{x_0} |v|\, dx.$$

Now if, along CD, $|v| < v_0$, we have

$$\lambda_3 - \lambda_2 \leqq 2Mx_0v_0 . \tag{2.18}$$

Otherwise, if $|v|$ exceeds v_0 along CD, we denote by P the point where $|v| = v_0$ for the first time between C and D and denote the coordinates of P by (x_4, v_0) and λ at P by λ_4'. Then, as above, integrating (2.13) from C to P,

$$\lambda_4' - \lambda_2 \leqq M \int_{x_4}^{x_0} |v|\, dx \leqq 2Mx_0v_0 .$$

Or, since $\lambda = \frac{1}{2}v^2 + G(x)$,

$$\tfrac{1}{2}v_0^2 - \tfrac{1}{2}v_2^2 \leqq 2Mx_0v_0 + G(x_0) - G(x_4)$$

$$\leqq 2Mx_0v_0 + G(x_0).$$

On the other hand, by (2.17), $\lambda_0 - \lambda_2 \geqq 5Mx_0v_0$, or

$$\tfrac{1}{2}v_0^2 - \tfrac{1}{2}v_2^2 \geqq 5Mx_0v_0 .$$

Thus $2Mx_0v_0 + G(x_0) \geqq 5Mx_0v_0$. Or

$$G(x_0) \geqq 3Mx_0v_0 \geqq 6Mx_0(G(x_1) - G(x_0))^{\frac{1}{2}}.$$

But since $G(x_1) - G(x_0) \geqq G^2(x_0)/M^2x_0^2$, this gives

$$G(x_0) \geqq 6G(x_0),$$

which is impossible. Thus $|v| < v_0$ and therefore (2.18) holds.

Combining (2.18) with (2.17), we have

$$\lambda_3 < \lambda_0 - 3Mx_0v_0 .$$

Since from D to E, $d\lambda/dt < 0$ along the integral curve, it follows that λ at E, which we shall denote by λ_4, is less than λ_3. Thus

$$\lambda_4 < \lambda_0 - 3Mx_0v_0 .$$

If we now proceed from E to F in much the same way as from C to D, it follows that λ at F is less than λ_0 which is the value of λ at A. Since $\lambda = \frac{1}{2}v^2 + G(x)$,

and $x = x_0$ at both A and F, it follows, therefore, that v at F is less than v at A. In other words, F lies below A as shown in Figure 2.

The above conclusion is valid if the integral curve intersects the line $x = -x_0$. If this does not occur then the integral curve cuts the x-axis at some point D' between 0 and $-x_0$. If the path from C to D' is treated in much the same way as that from C to D and that from D' to F in much the same way as from E to F, our conclusions will still follow.

We now denote by the region R the region which is bounded on the outside by $ABCDEFA$, and on the inside by the curve $\lambda(x, v) = \delta$, where δ is chosen so small that in the interior of $\lambda(x, v) = \delta$, $f(x, v) < 0$. Then no integral curve which starts in R will ever leave R as t increases. For in the first place such a curve will not cut across $ABCDEF$ since the latter is an integral curve and no two such curves intersect in a region free of singular points. In the second place, because $dx = v\,dt$, x increases with t in the upper half plane $v > 0$. Thus any integral curve in R and near the line segment AF will move away from AF to the right and therefore never intersect it. Hence the outer boundary of R will not be cut by any integral curve starting in R. Again, since along the inner boundary $f(x, v) < 0$, it follows from

$$\frac{d\lambda}{dt} = -f(x, v)v^2$$

that λ increases with t. Thus, integral curves in R near the inner boundary move outward from the inner boundary $\lambda(x, v) = \delta$ as t increases and, therefore, never intersect the inner boundary. Therefore the boundary of R is intersected by no integral curve which starts in R. By Theorem A of Bendixson, already referred to, this means that there is at least one closed integral curve which lies in R. This proves Theorem I.[7]

3. **Conditions for a unique solution.** Here we shall consider a condition which assures that the generalized equation for relaxation oscillations, (1.4), has, except for translations in t, only one periodic solution. This is equivalent to proving that there is at most one closed integral curve in the (x, v)-plane. From Theorem I we know of course that there must be at least one closed curve.

The method used here in proving that under certain conditions there is only one closed integral curve depends on the fact that two adjoining closed integral curves of (2.1) cannot both be stable. (A closed integral curve is stable if any integral curve starting sufficiently close to it spirals nearer and nearer to the closed integral curves as $t \to \infty$.)

To prove that two adjoining closed integral curves cannot both be stable we consider two stable closed integral curves I_1 and I_2 of (2.1). Let I_2 be interior to I_1. R is the region bounded outside by I_1 and inside by I_2. Obviously R

[7] Because of the form of the equations (2.1) it is also very easy to show how, directly without reference to Bendixson's theorem, a consequence of F lying below A in Figure 2 is that there must be at least one closed integral curve.

is free of singular points. We shall show that in the interior of R there must lie at least one closed integral curve I_3.

For, let us suppose that there is no closed integral curve in the interior of R. Let A, in Figure 3, be the point in which I_1 cuts the x-axis for $x > 0$ and B the point where I_2 cuts the x-axis for $x > 0$. Then every integral curve which starts on the line AB must, as $t \to \infty$, move closer and closer to I_1 or to I_2. This follows from the theorem of Bendixson used in §2, Theorem A, and from the fact that we have assumed that there are no closed curves in the interior of R. Consider a point A_1 on AB close to A. Then the integral curve starting at A_1 must, because of the stability of I_1, be asymptotic to I_1 and, therefore, as t increases, it must cut the positive x-axis at a sequence of points which moves steadily from A_1 to A

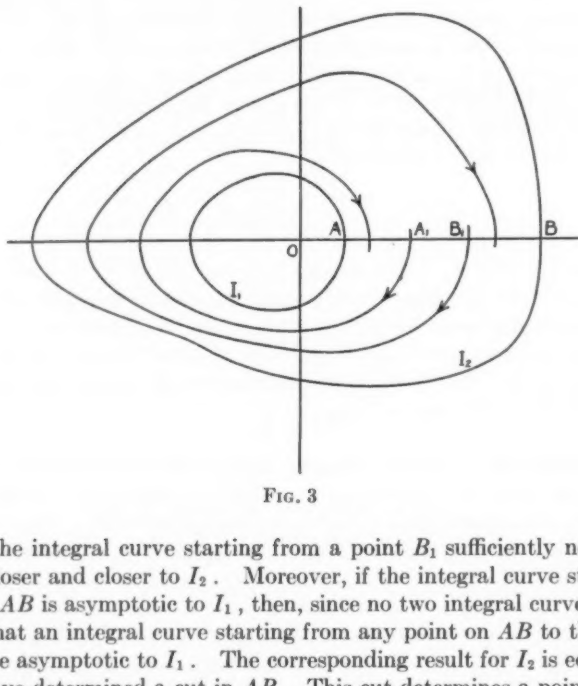

FIG. 3

Similarly, the integral curve starting from a point B_1 sufficiently near B must approach closer and closer to I_2. Moreover, if the integral curve starting at a point P on AB is asymptotic to I_1, then, since no two integral curves intersect, it follows that an integral curve starting from any point on AB to the left of P must also be asymptotic to I_1. The corresponding result for I_2 is equally true. Thus we have determined a cut in AB. This cut determines a point which we shall denote by C. The integral curve which starts at C and makes one turn about O must again intersect AB at C, for suppose it intersects to the right of C. Then by the continuity of integral curves with respect to changes in initial conditions, integral curves starting to the left of C, but very close to C, would also intersect the x-axis to the right of C after one turn and thus be asymptotic to I_2. But integral curves starting to the left of C must be asymptotic to I_1. Thus the integral curve starting at C cannot intersect AB to the right of C. In the

same way it cannot intersect AB to the left of C. Thus the integral curve at C is a closed curve I_3. This contradicts our assumption that I_1 and I_2 are two adjoining closed integral curves. Thus we have proved that two adjoining closed integral curves of (2.1) cannot both be stable.

We shall next show that under certain conditions every closed integral curve of (2.1) is stable. But this fact and the fact just demonstrated that there cannot be adjoining closed integral curves, both stable, means that there is at most one closed integral curve under these conditions. Thus theorems on the existence of a unique closed integral curve are reduced to theorems on stability.

Let us now consider the general equation for relaxation oscillations, (2.1), where $f(x, v)$ and $g(x)$ are subject to the conditions of Theorem I. As in §2 we consider the curves $\lambda(x, v) = c$. Further, we denote by R_1 the region in the (x, v)-plane where $f(x, v)$ is negative and by R_2 the region where $f(x, v)$ is positive. We shall denote that part of the curve $\lambda(x, v) = c$ which lies in R_1 by $R_1(c)$ and that part of $\lambda(x, v) = c$ which lies in R_2 by $R_2(c)$.

THEOREM II. *If the requirements of Theorem* I *are satisfied and if for every* c *the minimum of*

$$F(x, v) = \frac{1}{v^2} + \frac{1}{vf(x, v)} \frac{\partial f(x, v)}{\partial v}$$

on $R_2(c)$ *is positive and exceeds the maximum of* $F(x, v)$ *on* $R_1(c)$, *then the equations* (2.1) *possess a unique periodic solution.*[8]

As we have seen, the proof of this theorem will follow at once if we prove that under the conditions of this theorem every closed integral curve is stable. We shall therefore determine the condition for stability by the now classic method of Poincaré. We need only concern ourselves with the formal aspects of the condition here since the rigorous analysis is well known. Let us consider an integral curve. We denote the point at which this curve cuts the upper v-axis by v_0. We denote this curve by $v(x, v_0)$. Then let us next consider the integral curve which starts at the point $(0, v_0 + \delta v_0)$, where δv_0 is small. Then

$$v(x, v_0 + \delta v_0) = v(x, v_0) + \delta v_0 \frac{\partial v(x, v_0 + h\delta v_0)}{\partial v_0},$$

where $0 < h < 1$. Or, if we denote $v(x, v_0 + \delta v_0) - v(x, v_0)$ by Δv, then

$$\Delta v = \delta v_0 \frac{\partial v(x, v_0 + h\delta v_0)}{\partial v_0}. \tag{3.0}$$

If $v(x, v_0)$ is a closed integral curve, then for stability, it is necessary and sufficient that the integral curve starting at $(0, v_0 + \delta v_0)$ first again intersect the positive v-axis for increasing t at a point between v_0 and $v_0 + \delta v_0$. In other

[8] Actually the requirements of this theorem need only hold for $c < \Lambda$ where Λ is a constant such that it is known that there are no closed integral curves in the region $\lambda(x, v) \geqq \Lambda$. From the proof of Theorem I it is clear that Λ can be taken as λ_0.

words, that here $| \Delta v | < | \delta v_0 |$ for sufficiently small values of δv_0. Since $\partial v / \partial v_0$ is continuous, by (3.0) this is equivalent to $| \partial v(0, v_0)/\partial v_0 | < 1$ when the positive v-axis is intersected for the first time as t increases after starting from $(0, v_0)$ along the closed integral curve. From

$$\frac{dv}{dx} = -f(x, v) - \frac{g(x)}{v}$$

it follows for $v(x, v_0)$ that

$$\frac{d}{dx}\frac{\partial v}{\partial v_0} = -\frac{\partial f}{\partial v}\frac{\partial v}{\partial v_0} + \frac{g(x)}{v^2}\frac{\partial v}{\partial v_0}.$$

Thus

$$\frac{\dfrac{d}{dx}\left(\dfrac{\partial v}{\partial v_0}\right)}{\dfrac{\partial v}{\partial v_0}} = -\frac{\partial f}{\partial v} + \frac{g(x)}{v^2} = -\frac{\partial f}{\partial v} - \frac{1}{v}\frac{dv}{dx} - \frac{1}{v}f(x, v).$$

Or, integrating between two points A and B,

$$\log_e \left| \frac{\partial v}{\partial v_0} \right| \Bigg]_A^B = -\int_A^B \left[\frac{\partial f}{\partial v} + \frac{1}{v} f(x, v) \right] dx - \log_e | v | \Bigg]_A^B.$$

In particular, integrating once around a closed integral curve we take A and B as $(0, v_0)$. At A, the starting point on the positive v-axis, $\partial v/\partial v_0 = 1$. Moreover, $v_A = v_B = v_0$. Thus

$$\log_e \left| \frac{\partial v}{\partial v_0} \right|_B = -\int \left[\frac{\partial f}{\partial v} + \frac{1}{v} f(x, v) \right] dx,$$

where this last integral is extended around the closed integral curve. The condition that $| \partial v/\partial v_0 |_B < 1$ is clearly equivalent to

$$\int \left[\frac{\partial f}{\partial v} + \frac{1}{v} f(x, v) \right] dx > 0.$$

Since $dx = v\, dt$, this last integral can be written as

$$\int \left[v \frac{\partial f}{\partial v} + f(x, v) \right] dt > 0, \tag{3.1}$$

where the integral is taken around the closed integral curve. This, then, is a condition for stability.

We next concern ourselves with an integral which in certain applications is equivalent to an energy consideration. From (2.1) we have

$$v\phi[\lambda(x, v)] \left\{ \frac{dv}{dt} + vf(x, v) + g(x) \right\} = 0,$$

where $\phi(\lambda)$ is any integrable function of λ. This equation can be written as

$$\phi[\lambda(x, v)] \frac{d\lambda(x, v)}{dt} + v^2 \phi[\lambda(x, v)] f(x, v) = 0.$$

Integrating with respect to t around a closed integral curve, the first term drops out. Thus we have

$$\int v^2 \phi[\lambda(x, v)] f(x, v)\, dt = 0 \tag{3.2}$$

on a closed integral curve.

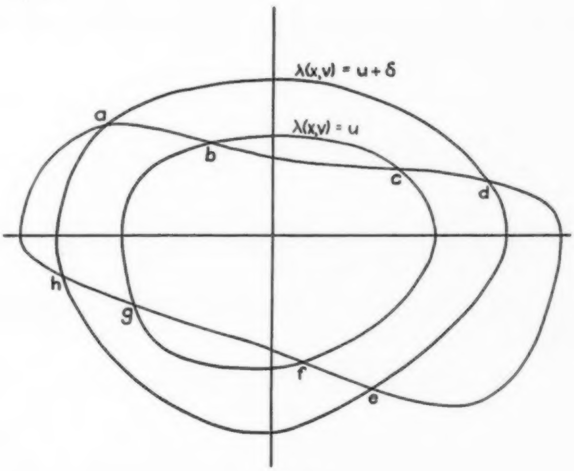

FIG. 4

Now let $\phi(\lambda) = 1/\delta$ when $u < \lambda < u + \delta$ and $\phi(\lambda) = 0$ otherwise. If $abcdefgh$ in Figure 4 represents a closed integral curve then

$$\int v^2 \phi[\lambda(x, v)] f(x, v)\, dt = \frac{1}{\delta}\left[\int_a^b + \int_c^d + \int_e^f + \int_g^h\right] v^2 f(x, v)\, dt.$$

Let

$$G_1(u) = \int_{R_1} v^2 \phi[\lambda(x, v)] f(x, v)\, dt,$$

$$G_2(u) = \int_{R_2} v^2 \phi[\lambda(x, v)] f(x, v)\, dt,$$

where the first integral is extended along that part of the closed integral curve in R_1 and the second, in R_2. From (3.2),

$$G_1(u) + G_2(u) = 0. \tag{3.3}$$

Now let

$$\lim_{\delta\to 0} G_1(u) = g_1(u), \qquad \lim_{\delta\to 0} G_2(u) = g_2(u).$$

Then from (3.3),

$$g_1(u) + g_2(u) = 0. \tag{3.4}$$

$$g_1(u) = \sum v^2 f(x, v) \left|\frac{dt}{d\lambda}\right|, \tag{3.5}$$

where the sum extends over the points where the closed integral curve is cut by $\lambda(x, v) = u$ in R_1. A similar result is true for $g_2(u)$. $dt/d\lambda$ is the reciprocal of $d\lambda/dt$ which is the velocity with which the curves $\lambda(x, v) = c$ are cut in phase space by the integral curve.

Next let

$$H_1(u) = \int_{R_1} \left[\frac{1}{vf}\frac{\partial f}{\partial v} + \frac{1}{v^2}\right] \phi[\lambda(x, v)] v^2 f(x, v)\, dt$$

and

$$H_2(u) = \int_{R_2} \left[\frac{1}{vf}\frac{\partial f}{\partial v} + \frac{1}{v^2}\right] \phi[\lambda(x, v)] v^2 f(x, v)\, dt,$$

where the first integral is extended along that part of the closed integral curve in R and the second, in R_2. Let

$$\lim_{\delta\to 0} H_1(u) = h_1(u), \qquad \lim_{\delta\to 0} H_2(u) = h_2(u).$$

Then clearly

$$\int \left[\frac{1}{vf}\frac{\partial f}{\partial v} + \frac{1}{v^2}\right] v^2 f(x, v)\, dt = \int_0^\infty [h_1(u) + h_2(u)]\, du. \tag{3.6}$$

Much as in (3.5)

$$h_1(u) = \sum \left[\frac{1}{vf}\frac{\partial f}{\partial v} + \frac{1}{v^2}\right] v^2 f(x, v) \left|\frac{dt}{d\lambda}\right|, \tag{3.7}$$

where the sum extends over the points where the closed integral curve is cut by $\lambda(x, v) = u$ in R_1 and a similar result holds for $h_2(u)$. Since $g_2(u) \geqq 0$ and $g_1(u) \leqq 0$ it follows from (3.5) and (3.7) that

$$h_1(u) \geqq \max_{R_1(u)} \left[\frac{1}{vf}\frac{\partial f}{\partial v} + \frac{1}{v^2}\right] g_1(u),$$

and similarly

$$h_2(u) \geqq \min_{R_2(u)} \left[\frac{1}{vf}\frac{\partial f}{\partial v} + \frac{1}{v^2}\right] g_2(u).$$

Thus, by the statement of Theorem II,

$$h_1(u) + h_2(u) \geqq \min_{R_2(u)} \left[\frac{1}{vf}\frac{\partial f}{\partial v} + \frac{1}{v^2}\right] (g_1(u) + g_2(u)).$$

Or, by (3.4),

$$h_1(u) + h_2(u) \geqq 0. \tag{3.8}$$

Now unless $\frac{1}{vf}\frac{\partial f}{\partial v} + \frac{1}{v^2}$ is constant on the curves $\lambda(x, v) = c$, the inequality sign in (3.8) must hold for almost all u. Thus

$$\int_0^\infty [h_1(u) + h_2(u)]\, du > 0$$

or, by (3.6),

$$\int \left[\frac{1}{vf}\frac{\partial f}{\partial v} + \frac{1}{v^2}\right] v^2 f(x, v)\, dt > 0.$$

But this is (3.1) which assures stability on every closed curve. This proves Theorem II except for the case where $(1/vf)(\partial f/\partial v) + 1/v^2$ is constant on the curves $\lambda(x, v) = c$. This last case is excluded by the hypothesis of Theorem I as we shall now show, for it implies that $(1/vf)(\partial f/\partial v) + 1/v^2$ is a function of $\lambda(x, v)$. That is,

$$\frac{1}{v}\frac{\partial}{\partial v} \log vf = \psi(\tfrac{1}{2}v^2 + G(x)).$$

Or, solving,

$$f(x, v) = \frac{K(x)}{v} \exp \int v\psi(\tfrac{1}{2}v^2 + G(x))\, dv,$$

where x is held constant during the integration in v. Clearly, no matter what function ψ is, this implies that $f(x, 0)$ is infinite for almost all x and, therefore, that (2.3) in the hypothesis of Theorem I cannot be fulfilled.

Theorem II has several useful corollaries which we shall now state and prove.

THEOREM III. *In the equations*

$$\frac{dv}{dt} + vf(x) + g(x) = 0, \qquad \frac{dx}{dt} = v,$$

$f(x)$ and $g(x)$ are differentiable functions. There exist an $x_1 > 0$ and an $x_2 > 0$ such that $f(x) < 0$ $(-x_1 < x < x_2)$ and $f(x) \geqq 0$ otherwise. Also $xg(x) > 0$ $(|x| > 0)$. Further, let

$$\int_0^{\pm\infty} g(x)\, dx = \int_0^\infty f(x)\, dx = \infty. \tag{3.9}$$

Also let

$$\int_0^x g(x)\,dx = G(x)$$

and suppose

$$G(-x_1) = G(x_2).$$

Then it follows that the equation has a unique periodic solution except for transitions in t.

Note that $f(x)$ need not be even nor $g(x)$ be odd, but that if this is the case then the requirement $G(-x_1) = G(x_2)$ is automatically satisfied.

Here the conditions of Theorem II become simply that

$$\min_{R_2(c)} \frac{1}{v^2} \geqq \max_{R_1(c)} \frac{1}{v^2}. \tag{3.10}$$

But R_1 is the strip $-x_1 \leqq x \leqq x_2$ and R_2 is $x \geqq x_2$ and $x \leqq -x_1$. Recall the fact that the curves $\lambda(x, v) = c$ have negative slope in the first and third quadrants of the (x, v)-plane and positive slope in the second and fourth. Also, since $G(-x_1) = G(x_2)$, it follows that $\frac{1}{2}v^2 + G(x) = c$ intersects the lines $x = -x_1$ and $x = x_2$ for the same positive and negative values of v, that is, let us say, at $\pm v_c$. But from this it follows at once that

$$\min_{R_2(c)} \frac{1}{v^2} = \frac{1}{v_c^2}, \qquad \max_{R_1(c)} \frac{1}{v^2} = \frac{1}{v_c^2}.$$

Thus (3.10) is satisfied and hence Theorem III is a consequence of Theorem II. The other conditions in Theorem III not used so far simply assure that the requirements of Theorem I are satisfied.

A more general corollary of Theorem II which includes Theorem III as a special case is

THEOREM IV. *Let the requirements of Theorem* I *be satisfied. Further, let there be an* $x_1 > 0$ *and an* $x_2 > 0$ *such that*

$$f(x, v) < 0 \qquad (-x_1 < x < x_2)$$

and $f(x, v) \geqq 0$ *otherwise. Let* $G(-x_1) = G(x_2)$ *and suppose*

$$v\frac{\partial f}{\partial v} \geqq 0.$$

Then (2.1) *possesses a unique periodic solution except for translations in* t.

The proof of this theorem goes in much the same way as the preceding one in so far as

$$\min_{R_2(c)} \frac{1}{v^2} \geqq \max_{R_1(c)} \frac{1}{v^2} \tag{3.11}$$

goes. Further, by $v(\partial f/\partial v) \geqq 0$,

$$\frac{1}{f(x, v)} \frac{1}{v} \frac{\partial f}{\partial v} \geqq 0$$

in R_2 since here $f \geqq 0$, and

$$\frac{1}{fv} \frac{\partial f}{\partial v} \leqq 0$$

in R_1, where $f \leqq 0$. From this and (3.11) it follows at once that

$$\min_{R_2(c)} \left[\frac{1}{v^2} + \frac{1}{fv} \frac{\partial f}{\partial v}\right] \geqq \max_{R_1(c)} \left[\frac{1}{v^2} + \frac{1}{fv} \frac{\partial f}{\partial v}\right]$$

and thus Theorem IV is a consequence of Theorem II.

4. **Liénard's method.** Here we shall consider the equation

$$\ddot{x} + f(x)\dot{x} + g(x) = 0, \tag{4.0}$$

where $f(x)$ is an even function such that for the odd function

$$F(x) = \int_0^x f(x)\, dx$$

there exists an x_0 with $F(x) < 0$ for $0 < x < x_0$, and $F(x) > 0$ and monotonically increasing for $x > x_0$. Moreover, $g(x)$ is an odd differentiable function such that $g(x) > 0$ for $x > 0$. We further assume that

$$\int_0^\infty f(x)\, dx = \int_0^\infty g(x)\, dx = \infty \tag{4.1}$$

although milder conditions of the type given in Theorem I would suffice. *Under these conditions we shall show that* (4.0) *possesses a unique periodic solution.*

For the case $g(x) = x$ this result has been demonstrated by Liénard. Here we shall modify the proof of Liénard so that it applies to (4.0). The result here is more inclusive than Theorem III of §3 in so far as the requirements on $F(x)$ go, but more restrictive in requiring that $f(x)$ be even and $g(x)$ odd.

As before, we set $\dot{x} = v$. Then (4.0) becomes

$$\frac{dv}{dx} + f(x) + \frac{g(x)}{v} = 0. \tag{4.2}$$

We now introduce $y = v + F(x)$. Then (4.2) can be written as

$$\frac{dy}{dx} + \frac{g(x)}{y - F(x)} = 0. \tag{4.3}$$

Clearly, a unique periodic solution of (4.0) is equivalent to a unique closed integral curve for (4.2) which in turn is equivalent to a unique closed integral curve for (4.3). Moreover, since (4.3) remains unchanged if (x, y) is replaced

by $(-x, -y)$ it follows that if a closed integral curve passes through $(0, y_0)$ it must also pass through $(0, -y_0)$. For if this were not the case the reflection in the origin, that is, replacing of (x, y) by $(-x, -y)$, would give rise to another closed integral curve which intersects the first one. But two integral curves cannot intersect, except at the origin, and thus a closed integral curve starting at $(0, y_0)$ passes through $(0, -y_0)$. In fact, what we have really shown is that a closed integral curve is symmetric with respect to the origin. Conversely, any integral curve starting at $(0, y_0)$ and passing through $(0, -y_0)$ must be closed since on leaving $(0, -y_0)$ it must follow the reflection in the origin of the path from $(0, y_0)$ to $(0, -y_0)$. Thus to find a closed integral curve of (4.3) is equivalent to finding an integral curve with positive and negative y intercepts equal.

In showing that there is only one integral curve with positive y intercept equal to its negative y intercept, we shall study the change in intercepts of the integral curves by studying how these integral curves cut across the curves $\lambda(x, y) = c$, where

$$\lambda(x, y) = \tfrac{1}{2}y^2 + G(x).$$

As in the previous articles

$$G(x) = \int_0^x g(x)\, dx.$$

Clearly $\lambda(x, y)$ is symmetric with respect to changes in sign in both x and y. Thus, if $\lambda(0, y)$ is the same for an integral curve for both positive and negative y, then the integral curve is closed.

Looking at (4.3), it follows at once that, for $x > 0$, integral curves have negative slope where $y > F(x)$ and positive slope where $y < F(x)$. The slope is infinite where $y = F(x)$. Thus in Figure 5 ACB, $A'C'B'$, and $A''C''B''$ are all integral curves. The equation (4.3) can be written as

$$ydy + g(x)dx = F(x)dy$$

or as

$$d\lambda(x, y) = F(x)dy. \tag{4.4}$$

We now consider a section of an integral curve ACB such that C, the intersection of the curve with $y = F(x)$, falls in the strip $0 < x < x_0$. In this strip, $F(x) < 0$. Moreover, from A to B, $dy < 0$. Thus $F(x)dy > 0$ and from (4.4)

$$\int_A^B d\lambda(x, y) > 0$$

or, in other words, $\lambda_B - \lambda_A > 0$. Thus, $\overline{OB} > \overline{OA}$.

Next we consider integral curves which intersect $y = F(x)$ to the right of $x = x_0$. $A'B'C'$ and $A''B''C''$ are two such curves. From (4.3) and (4.4)

$$d\lambda(x, y) = \frac{-F(x)g(x)}{y - F(x)}\, dx. \tag{4.5}$$

Since $-F(x) > 0$ for $0 < x < x_0$ and since $y - F(x)$ is greater along $A''G$ than along $A'E$, it follows from this equation that

$$\int_{A''}^{G} d\lambda(x, y) < \int_{A'}^{E} d\lambda(x, y),$$

each integral being taken along the proper integral curve. In other words,

$$\lambda_G - \lambda_{A''} < \lambda_E - \lambda_{A'} . \tag{4.6}$$

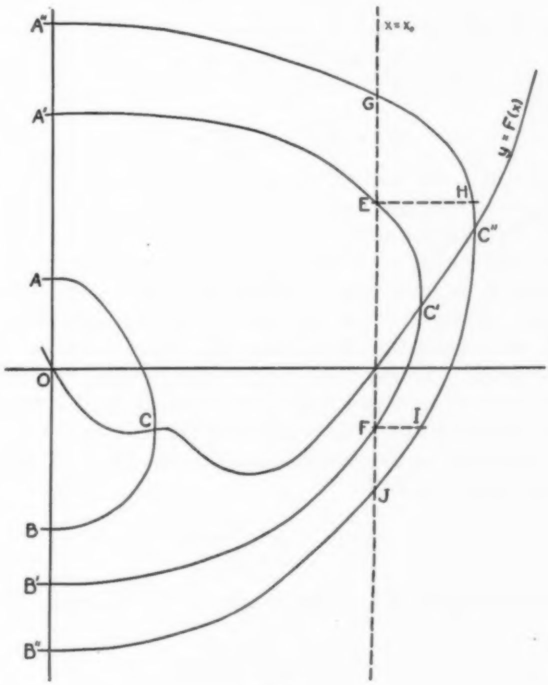

FIG. 5

From $d\lambda(x, y) = F(x)dy$, it follows since $F(x) > 0$ along GH that $d\lambda(x, y) < 0$. Thus

$$\lambda_H - \lambda_G < 0. \tag{4.7}$$

Since for the same y, $F(x)$ along HI exceeds $F(x)$ along EF, it follows from $d\lambda(x, y) = F(x)dy$ that

$$\int_{H}^{I} d\lambda(x, y) < \int_{E}^{F} d\lambda(x, y)$$

or that

$$\lambda_I - \lambda_H < \lambda_F - \lambda_E . \tag{4.8}$$

Just as along GH, so also along IJ it follows that

$$\lambda_J - \lambda_I < 0. \tag{4.9}$$

Similarly just as (4.6) is obtained we get

$$\lambda_{B''} - \lambda_J < \lambda_{B'} - \lambda_F . \tag{4.10}$$

Adding (4.6) through (4.10), we get

$$\lambda_{B''} - \lambda_{A''} < \lambda_{B'} - \lambda_{A'} .$$

In other words,

$$\overline{OB}'' - \overline{OA}'' \leqq \overline{OB}' - \overline{OA}'.$$

In other words, $\overline{OB} - \overline{OA} > 0$ so long as c lies in $0 < x \leqq x_0$. When C moves outward along $y = F(x)$ and cuts $y = F(x)$ for $x > x_0$, then $\overline{OB} - \overline{OA}$ is a monotonically decreasing function. Clearly, then, $\overline{OB}$ can equal $\overline{OA}$ at most once, meaning there is at most one closed integral curve. From our general theory, Theorem I, we know there is at least one closed integral curve and thus we are through. However, since the proof so far has been quite elementary, we shall give an independent proof that $\overline{OB} - \overline{OA}$ becomes negative as A moves up the y-axis. To see this we have only to observe that the increase in λ in going from A'' to G and from J to B'' is monotonically decreasing as A'' moves up the y-axis and is therefore bounded. From G to J, λ decreases. Consider the intercepts of the integral curves on the line $x = 2x_0$ as A'' moves up the y-axis. Since by (4.3)

$$dy = -\frac{g(x)}{y - F(x)}\, dx,$$

it follows that starting at A'', in the interval $0 < x < 2x_0$,

$$|\, dy \,| \leqq \frac{g(x)\, dx}{|\, y - F(2x_0)\, |}.$$

Or if y_0 denotes $y(2x_0)$, we have on integrating

$$\frac{\overline{OA}''^2}{2} - \frac{y_0^2}{2} - (\overline{OA}'' - y_0)F(2x_0) \leqq G(2x_0).$$

Or

$$(\overline{OA}'' - y_0)\left[\frac{\overline{OA}'' + y_0}{2} - F(2x_0)\right] \leqq G(2x_0).$$

Or

$$\overline{OA}'' - y_0 \leqq \frac{2G(2x_0)}{\overline{OA}'' - 2F(2x_0)}.$$

Thus, as $\overline{OA}'' \to \infty$, $y_0 \to \overline{OA}''$. Since no two integral curves intersect, it follows that the distance between the intercepts on $x = 2x_0$ of an integral curve for which $\overline{OA}'' \to \infty$ is also tending toward infinity. But $d\lambda = F(x)dy$. Since, for $x > 2x_0$, $F(x) > F(2x_0)$, $d\lambda < F(2x_0)dy$. Integrating between the intercepts on $x = 2x_0$,

$$\int d\lambda < -F(2x_0)D,$$

where D is the distance between the intercepts. Since $D \to \infty$, it follows that the decrease in λ in going from G to J can be made arbitrarily large. On the other hand, the increase from A'' to G and from J to B'' is, as we have seen, bounded. Thus, for large $\overline{OA}$, $\lambda_B - \lambda_A < 0$ or $\overline{OB} - \overline{OA} < 0$. This completes the proof.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

THE DIVERGENCE OF NON-HARMONIC GAP SERIES

By Philip Hartman

It was recently shown[1] that if $\lambda_0, \lambda_1, \cdots$ is a sequence of positive real numbers satisfying the gap condition

$$\frac{\lambda_k}{\lambda_{k-1}} > q > 1 \qquad (k = 1, 2, \cdots), \tag{1}$$

then the convergence of the series

$$\sum_{k=0}^{\infty} | a_k |^2 \tag{2}$$

implies the convergence of

$$\sum_{k=0}^{\infty} a_k e^{i\lambda_k t} \tag{3}$$

for almost all t, $-\infty < t < +\infty$, while if (1) is modified so that "$q > 1$" is replaced by "$q > \frac{1}{2}(5^{\frac{1}{2}} + 1)$", then the divergence of (2) implies the divergence of (3) for almost all t, $-\infty < t < +\infty$. The object of this note is to show that *the condition* (1), without any modification, *and the divergence of* (2) *implies the divergence of the series* (3) *for almost all* t, $-\infty < t < +\infty$.

In order to prove this statement, let $\sigma(E)$ denote the completely additive measure on the t-axis which has the non-negative density[2]

$$\frac{(1 - \cos t)}{\pi t^2} \qquad (-\infty < t < +\infty) \tag{4}$$

so that, if E is a measurable set,

$$\sigma(E) = \int_E \frac{1 - \cos t}{\pi t^2}\, dt. \tag{5}$$

Obviously, for any measurable set E,

$$0 \leqq \sigma(E) \leqq 1. \tag{6}$$

The Fourier-Stieltjes transform of this σ-measure,

$$\int_{-\infty}^{+\infty} e^{i\lambda t}\, d\sigma(t) = \max(1 - |\lambda|, 0), \tag{7}$$

vanishes for all $|\lambda| \geqq 1$.

Received January 5, 1942.

[1] M. Kac, *Convergence and divergence of non-harmonic gap series*, Duke Mathematical Journal, vol. 8(1941), pp. 541–545.

[2] The introduction of this measure function avoids the awkward construction, used in loc. cit., see footnote 1, of a measure whose Fourier-Stieltjes transform vanishes on a particular sequence of points. It also makes it possible to use, with only the slightest of modifications, the method applied by A. Zygmund, *Trigonometrical Series*, Warsaw, 1935, pp. 120–122, in the case that the frequencies λ_k are integers.

Let E be any measurable set and let m, n $(n > m)$ be a pair of positive integers. Then

$$(8)\qquad \int_E \left| \sum_{k=m}^{n} a_k e^{i\lambda_k t} \right|^2 d\sigma(t) = \sigma(E) \sum_{k=m}^{n} | a_k |^2 + \sum_{k=m}^{n} \sum_{\substack{k=m \\ k\neq j}}^{n} a_k \bar{a}_j \int_E e^{i(\lambda_k - \lambda_j)t} d\sigma(t).$$

By Schwarz's inequality, the absolute value of the last term is majorized by

$$(9)\qquad \left(\sum_{k=m}^{n} \sum_{\substack{k=m \\ k\neq j}}^{n} | a_k \bar{a}_j |^2 \right)^{\frac{1}{2}} \left(\sum_{k=m}^{n} \sum_{\substack{k=m \\ k\neq j}}^{n} \left| \int_E e^{i(\lambda_k - \lambda_j)t} d\sigma(t) \right|^2 \right)^{\frac{1}{2}}.$$

To appraise the last expression, it is necessary to consider the structure of the set of numbers $\lambda_k - \lambda_j$ $(k > j; j, k = 0, 1, \cdots)$. First, there exists a positive number δ such that $\lambda_k - \lambda_j > \delta$ for all $k > j$. In fact, by (1), $\lambda_k - \lambda_j \geqq \lambda_k - \lambda_{k-1} > \lambda_k(1 - q^{-1}) \geqq \lambda_0(1 - q^{-1})$ so that δ may be taken to be $\lambda_0(1 - q^{-1})$. Secondly, the number of numbers $\lambda_k - \lambda_j$, $k > j$, in any interval $c \leqq \lambda \leqq c + 1$, $c > \delta$, is bounded. For if $c \leqq \lambda_k - \lambda_j \leqq c + 1$, then $c \leqq \lambda_k$ and $c + 1 \geqq \lambda_k(1 - q^{-1})$. In virtue of (1) the number of integers k such that $q(c + 1)/(q - 1) \geqq \lambda_k \geqq c$ is at most $\{\log (c + 1)/c + \log q/(q - 1)\}/\log q$, which is bounded for $\delta \leqq c < +\infty$. From these two properties of the numbers $\lambda_k - \lambda_j$, $k > j$, it follows that the set of numbers $\lambda_k - \lambda_j$ $(k \neq j; j, k = 0, 1, \cdots)$ can be divided into a finite number of sequences, say N sequences, such that the absolute value of the difference of any two numbers in the same sequence exceeds 1. The identity (7) implies that each of the corresponding sequences of functions $e^{i(\lambda_k - \lambda_j)t}$ forms an orthogonal sequence on $-\infty < t < +\infty$ with respect to the σ-measure. Hence, by the Bessel inequality, the series

$$\sum_{k=0}^{\infty} \sum_{\substack{j=0 \\ k\neq j}}^{\infty} \left| \int_E e^{i(\lambda_k - \lambda_j)t} d\sigma(t) \right|^2$$

converges (and has a sum which does not exceed $N\, \sigma(E)$).

Suppose that the integer m is chosen so large that

$$\sum_{k=m}^{\infty} \sum_{\substack{j=m \\ k\neq j}}^{\infty} \left| \int_E e^{i(\lambda_k - \lambda_j)t} d\sigma(t) \right|^2 \leqq (\tfrac{1}{2}\sigma(E))^2.$$

Then, from (8) and (9),

$$\int_E \left| \sum_{k=m}^{n} a_k e^{i\lambda_k t} \right|^2 d\sigma(t) \geqq \tfrac{1}{2}\sigma(E) \sum_{k=m}^{n} | a_k |^2.$$

This inequality and the finiteness (6) of the σ-measure implies that if the series (2) diverges and if E is a measurable set such that (3) converges for all points t on E, then $\sigma(E) = 0$. But since the density (4) of σ vanishes for only an enumerable set of t-values, it follows from $\sigma(E) = 0$ that the Lebesgue measure of E is 0. This completes the proof of the italicized statement.

Fort Bragg, N. C.

INFLUENCE OF THE SIGNS OF THE DERIVATIVES OF A FUNCTION ON ITS ANALYTIC CHARACTER

By R. P. Boas, Jr. and G. Pólya

1. **Introduction.** In what follows, $f(x)$ denotes a real-valued function defined and of class C^∞ in $[-1, 1]$, i.e., possessing derivatives of all orders in the closed interval $-1 \leqq x \leqq 1$.[1]

Serge Bernstein investigated the analytic nature of functions whose derivatives are each of constant sign in $[-1, 1]$. That such a function is necessarily analytic in $(-1, 1)$ is contained as a very special case in one of his earlier theorems.[2] But he observed also that the signs of the derivatives have a certain influence.[3] If no derivative of $f(x)$ vanishes in $(-1, 1)$, $|f^{(n)}(x)|$ is either steadily increasing or steadily decreasing; we have the first or the second case according as $f^{(n)}(x)f^{(n+1)}(x) > 0$ or $f^{(n)}(x)f^{(n+1)}(x) < 0$ in $(-1, 1)$ (consider the derivative of $[f^{(n)}(x)]^2$). We say that $f^{(m)}(x)$ and $f^{(n)}(x)$ (where $m < n$) belong to the same "block" if $|f^{(m)}(x)|, |f^{(m+1)}(x)|, \cdots, |f^{(n-1)}(x)|, |f^{(n)}(x)|$ all vary in the same sense, i.e., all increase or all decrease. Thus, $f^{(n)}(x)$ and $f^{(n+1)}(x)$ belong to different blocks if and only if

$$f^{(n)}(x)f^{(n+2)}(x) < 0.$$

Let $\lambda_1, \lambda_2, \lambda_3, \cdots$ denote the lengths of the successive blocks into which the sequence $f(x), f'(x), f''(x), \cdots$ is decomposed; we assume here that no block has infinite length, and that, therefore, there is an infinity of blocks. Bernstein found the remarkable result that the lengths of the blocks influence the analytic nature of the function. Roughly stated, the analytic nature of $f(x)$ is simpler if the blocks are shorter. E.g., if the sequence $\lambda_1, \lambda_2, \cdots$ is bounded, $f(x)$ is (or, more precisely, coincides in $[-1, 1]$ with) an entire function of exponential type, i.e., an entire function whose growth does not exceed order one and finite type.

Received January 5, 1942.

[1] We write $[a, b]$ for the closed interval $a \leqq x \leqq b$, and (a, b) for the open interval $a < x < b$. The conventions about $f(x)$ do not apply to section 3.

[2] S. Bernstein, (a) *Sur la définition et les propriétés des fonctions analytiques d'une variable réelle*, Mathematische Annalen, vol. 75(1914), pp. 449–468, (b) *Leçons sur les propriétés extrémales et la meilleure approximation des fonctions analytiques d'une variable réelle*, Paris, 1926; see especially pp. 196–197. Another proof has been given by R. P. Boas, *Functions with positive derivatives*, this Journal, vol. 8(1941), pp. 163–172.

[3] S. Bernstein, (a) *On certain properties of regularly monotonic functions* (in Russian), Soobshcheniya Kharkovskogo Matematicheskogo Obshchestva (Communications de la Société Mathématique de Kharkow), (4), vol. 2(1928), pp. 1–11, (b) *Sur les fonctions régulièrement monotones*, Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, Paris, vol. 186(1928), pp. 1266–1269, (c) Same title, Atti del Congresso Internazionale dei Matematici, Bologna, 1928, vol. 2(1930), pp. 267–275.

Observe that since $f(x)$ is analytic in $(-1, 1)$ and not a polynomial, if no derivative changes sign in $(-1, 1)$, then no derivative can vanish there.

D. V. Widder found recently that $f(x)$ is necessarily an entire function of exponential type if[4]

$$(-1)^n f^{(2n)}(x) \geqq 0 \qquad (-1 \leqq x \leqq 1; n = 0, 1, 2, \cdots).$$

Widder's condition would imply, *if* we knew that no derivative of $f(x)$ vanishes in $(-1, 1)$, that $f^{(2n)}(x)$ and $f^{(2n+1)}(x)$ do not belong to the same block, and that, therefore, no block has a length greater than two. But, in fact, Widder's condition does not say anything about the non-vanishing of the derivatives of odd order in $(-1, 1)$, and therefore Widder's theorem is not contained in Bernstein's results.

The following theorem contains both Bernstein's and Widder's results.[5]

THEOREM 1. *Let $\{n_k\}$ and $\{q_k\}$ be sequences of positive integers, $\{n_k\}$ strictly increasing. Let $f(x)$ be real-valued and of class C^∞ in $[-1, 1]$. For $k = 1, 2, \cdots$, let $f^{(n_k)}(x)$ and $f^{(n_k+2q_k)}(x)$ not change sign in $[-1, 1]$, and let*

$$f^{(n_k)}(x) f^{(n_k+2q_k)}(x) \leqq 0.$$

(I) *If $n_k - n_{k-1} = O(1)$ and $q_k = O(1)$, then $f(x)$ coincides in $[-1, 1]$ with an entire function of growth not exceeding order one and finite type.*

(II) *If $n_k - n_{k-1} = O(n_k^\delta)$, $q_k = O(n_k^\delta)$, and $q_1 + q_2 + \cdots + q_k = O(n_k)$, where δ is fixed, $0 < \delta < 1$, then $f(x)$ coincides in $[-1, 1]$ with an entire function of finite order not exceeding $1/(1 - \delta)$.*

(III) *If $n_k - n_{k-1} = o(n_k)$, $q_k = o(n_k)$, and $q_1 + q_2 + \cdots + q_k = O(n_k)$, then $f(x)$ coincides in $[-1, 1]$ with an entire function.*

In order to apply this theorem to Bernstein's case, in which no derivative vanishes in $(-1, 1)$, let $f^{(n_k)}(x)$ denote the *last* derivative belonging to the k-th block, so that

$$f^{(n_k)}(x) f^{(n_k+2)}(x) < 0,$$

$$\lambda_1 = n_1, \qquad \lambda_k = n_k - n_{k-1} \qquad \text{for } k > 1.$$

Put $q_k = 1$; then the hypothesis of Theorem 1 is fulfilled, and the specialization performed gives us exactly Bernstein's results.[6] On the other hand, assuming that $2q_k = n_{k+1} - n_k$ and that n_1 is even, we obtain from Theorem 1 the following direct generalization of Widder's result.

THEOREM 2. *Let $\{n_k\}$ be a strictly increasing sequence of positive even integers. Let $f(x)$ be real-valued and of class C^∞ in $[-1, 1]$, and let*

$$(-1)^k f^{(n_k)}(x) \geqq 0 \qquad (k = 1, 2, \cdots).$$

[4] D. V. Widder, *Functions whose even derivatives have a prescribed sign*, Proceedings of the National Academy of Sciences, vol. 26(1940), pp. 657–659.

[5] The main results of the present paper were stated in a joint note by the authors (*Generalizations of completely convex functions*, Proceedings of the National Academy of Sciences, vol. 27(1941), pp. 323–325), where previous contributions of both authors to the problem are quoted.

[6] See footnote 3, (a), pp. 4–5.

(I) *If* $n_k - n_{k-1} = O(1)$, $f(x)$ *coincides in* $[-1, 1]$ *with an entire function of growth not exceeding order one and finite type.*

(II) *If* $n_k - n_{k-1} = O(n_k^\delta)$, $0 < \delta < 1$, $f(x)$ *coincides in* $[-1, 1]$ *with an entire function of finite order not exceeding* $1/(1 - \delta)$.

(III) *If* $n_k - n_{k-1} = o(n_k)$, $f(x)$ *coincides in* $[-1, 1]$ *with an entire function.*

We cannot prove Theorem 1 by Bernstein's method because, under its hypothesis, many derivatives of $f(x)$ may change sign in $[-1, 1]$ (see (IV) in §6); and we cannot prove it by Widder's method, because the sequence $n_1, n_2, n_3, \cdots$ may be much more irregular than the extremely regular special sequence 2, 4, 6, $\cdots$. Our proof (see §5) is based on an inequality for derivatives, in which the constant sign of a derivative and the evenness of the difference between its order and that of another derivative are aptly combined (see Lemma 6 in §4).

We thought it appropriate to include proofs of the main facts on which Lemma 6 is based (see Lemmas 1 to 5 in §§2 and 3). These facts constitute an important part of the technique of dealing with the derivatives of real functions, and our proofs are somewhat simpler than previous proofs.

The last section of the paper (§6) is devoted to additional remarks and to the construction of examples showing that the results stated in Theorem 1 are fairly sharp.

2. **Derivatives of polynomials.** We start from the following well-known and easily proved fact.[7]

LEMMA 1. *Assume that* $P(x)$ *is a polynomial of degree not exceeding* n, M *is the maximum of* $|P(x)|$ *in* $[-1, 1]$, z *is a point of the complex plane,* a *is the larger and* b *is the smaller semi-axis of the ellipse with foci at the points* $+1$ *and* -1 *of the complex plane, passing through* z.

Then

$$|P(z)| \leqq M(a + b)^n.$$

We use Lemma 1 to prove the following[8]

LEMMA 2. *Under the hypothesis of Lemma* 1,

[7] This theorem is due to S. Bernstein and plays an important rôle in his work. For proof and further references see, e.g., G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, Berlin, 1925, vol. 1, section III, problem 270, pp. 137, 320.

[8] The result of Lemma 2 is less exact but the proof is much simpler than that of W. Markoff, *Über Polynome, die in einem gegebenen Intervalle möglichst wenig von Null abweichen* (in Russian), St. Petersburg, 1892; German abridgement in Mathematische Annalen, vol. 77(1916), pp. 213–258. The method of proof is that of P. Montel, *Sur les polynomes d'approximation*, Bulletin de la Société Mathématique de France, vol. 46(1918), pp. 151–192, 157ff. See for the whole question the address of A. C. Schaeffer, *Inequalities of A. Markoff and S. Bernstein for polynomials and related functions*, Bulletin of the American Mathematical Society, vol. 47(1941), pp. 565–579.

$$(1) \qquad | P^{(k)}(0) | \leqq 3k^{\frac{1}{2}} n^k M,$$

$$(1^*) \qquad | P^{(k)}(x) | \leqq \frac{9k}{k!\,2^k} n^{2k} M$$

for $k = 1, 2, \cdots, n - 1$ *and* $-1 \leqq x \leqq 1$.

It is easy to see that the ellipse with foci 1 and -1 and semi-axes a and b contains the circle with center 0 and radius b, and also any circle whose center is a point x of $[-1, 1]$ and whose radius is $a - 1$. Hence, by Lemma 1 and Cauchy's estimate for the absolute value of the k-th derivative,

$$(2) \qquad | P^{(k)}(0) | \leqq k! \frac{(a + b)^n}{b^k} M,$$

$$(2^*) \qquad | P^{(k)}(x) | \leqq k! \frac{(a + b)^n}{(a - 1)^k} M.$$

We have still to choose the ellipse; we have, of course, the condition that

$$a^2 - b^2 = 1.$$

Under this condition, we seek the minimum of the right side of (2). The usual procedure leads us to the values

$$a = \frac{n}{(n^2 - k^2)^{\frac{1}{2}}}, \qquad b = \frac{k}{(n^2 - k^2)^{\frac{1}{2}}}, \qquad a + b = \left(\frac{n + k}{n - k}\right)^{\frac{1}{2}};$$

substituting these values into the right side of (2), we obtain

$$(3) \qquad | P^{(k)}(0) | \leqq \frac{k!}{k^k} \frac{(n + k)^{(n+k)/2}}{(n - k)^{(n-k)/2}} M.$$

Minimizing the right side of (2*) yields

$$a = \frac{n^2 + k^2}{n^2 - k^2}, \qquad b = \frac{2nk}{n^2 - k^2}, \qquad a + b = \frac{n + k}{n - k},$$

and these values change (2*) into

$$(3^*) \qquad | P^{(k)}(x) | \leqq \frac{1}{k!\,2^k} \left\{\frac{k!}{k^k} \frac{(n + k)^{(n+k)/2}}{(n - k)^{(n-k)/2}}\right\}^2 M.$$

Comparison of (3) and (3*) with each other and with (1) and (1*) shows that, from this point on, it is sufficient to consider (3). Now

$$(4) \qquad \frac{(n + k)^{(n+k)/2}}{(n - k)^{(n-k)/2}} = n^k e^{k\varphi(k/n)},$$

where we use the abbreviation

$$[(1 + x) \log (1 + x) - (1 - x) \log (1 - x)]/(2x) = \varphi(x) \qquad (0 < x < 1);$$

$\varphi(x)$ is defined by continuity in $[0, 1]$. We observe that in $(0, 1)$

$$\varphi'(x) = \frac{1}{x^2}\left[x - \tfrac{1}{2}\log\frac{1+x}{1-x}\right] = \frac{1}{x^2}\left[-\frac{x^3}{3} - \frac{x^5}{5} - \cdots\right] < 0,$$

and therefore that, for $0 < x < 1$,

$$\varphi(x) < \varphi(0) = 1.$$

Combining this with (4), (3), (3*), and the well-known inequality

$$k! < (2\pi)^{\frac{1}{2}} k^{k+\frac{1}{2}} e^{-k+1/(12k)} < 3k^{\frac{1}{2}} k^k e^{-k},$$

we obtain both (1) and (1*).

3. **Derivatives of real functions.** We use Lemma 2 to prove the following lemma.[9]

LEMMA 3. *Let the function $f(x)$ be defined and possess an n-th derivative in* $[-l, l]$,[10] *and let it satisfy in this interval the conditions*

$$|f(x)| \leqq M_0, \qquad |f^{(n)}(x)| \leqq M_n. \tag{5}$$

Define

$$M'_n = \max(M_n, n!M_0 l^{-n}), \tag{6}$$

$$M^*_n = \max(M_n, n!M_0(2l)^{-n}). \tag{6*}$$

Then

$$|f^{(k)}(0)| \leqq 6k^{\frac{1}{2}} e^k M_0^{1-k/n} M_n'^{k/n}, \tag{7}$$

[9] Lemma 3 is essentially equivalent to two theorems due to A. Gorny, *Contribution à l'étude des fonctions dérivables d'une variable réelle*, Acta Mathematica, vol. 71(1939), pp. 317–358. Our proof differs in two points from Gorny's proof. (I) Instead of the best polynomial approximation to $f(x)$, we use the approximation given by Taylor's formula; this method was indicated (before Gorny) by O. Ore, *On functions with bounded derivatives*, Transactions of the American Mathematical Society, vol. 43(1938), pp. 321–326. (II) Instead of the best estimate for the k-th derivative of a polynomial, we use the approximate estimate of Lemma 2; the possibility of such a variant was hinted by Gorny (op. cit., p. 321, footnote).

Essentially equivalent theorems have been announced by H. Cartan. See H. Cartan and S. Mandelbrojt, *Solution du problème d'équivalence des classes de fonctions indéfiniment dérivables*, Acta Mathematica, vol. 72(1940), pp. 31–49. More precise results, applying only to the interval $(-\infty, \infty)$, are given by A. Kolmogoroff, *On inequalities between upper bounds of consecutive derivatives of an arbitrary function defined on an infinite interval* (in Russian; English summary), Uchenye Zapiski Moskovskogo Gosudarstvennogo Universiteta, Matematika, vol. 30(1939), pp. 3–16; the results of this paper are also stated in *Une généralisation de l'inégalité de M. J. Hadamard entre les bornes supérieures des dérivées successives d'une fonction*, Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, Paris, vol. 207(1938), pp. 764–765.

(Added in proof). Cartan's results have appeared in a publication which reached this country while this paper was in the press (*Sur les classes de fonctions définies par des inégalités portant sur leurs dérivées successives*, Actualités Scientifiques et Industrielles, no. 867, Paris, 1940). Cartan's proof, like ours, uses Taylor's formula.

[10] In section 3 the function $f(x)$ need not have derivatives of all orders.

$$|f^{(k)}(x)| \leqq \frac{18k}{k!} e^k n^k M_0^{1-k/n} M_n^{*k/n} \tag{7*}$$

for $k = 1, 2, \cdots, n$ *and* $-l \leqq x \leqq l$.

We shall prove in detail (7), which we must use later, and we shall indicate the points where the proof of (7*) differs from that of (7). We assume in the proofs that $l = 1$ (the general case is reduced to this special case by consideration of $f(lx)$). We divide the proofs into two parts.

(I) We consider the polynomial

$$\begin{aligned} P(x) &= f(0) + \frac{xf'(0)}{1!} + \cdots + \frac{x^{n-1}f^{(n-1)}(0)}{n!} \\ &= f(x) - \frac{x^n f^{(n)}(\theta x)}{n!}, \end{aligned} \tag{8}$$

where $0 < \theta < 1$ (we use Lagrange's form of the remainder of Taylor's series). By (5) and (8),

$$|P(x)| \leqq M_0 + \frac{M_n}{n!}$$

in $[-1, 1]$; hence it follows by Lemma 2 that

$$|f^{(k)}(0)| = |P^{(k)}(0)| \leqq 3k^{\frac{1}{2}} n^k \left(M_0 + \frac{M_n}{n!}\right) \tag{9}$$

for $k = 1, 2, \cdots, n - 1$. From (9), (7) can be obtained very quickly (see below).

(I*) In order to prove (7*), we consider the polynomial in x

$$\begin{aligned} P(x) &= f(\xi) + \frac{(x-\xi)f'(\xi)}{1!} + \cdots + \frac{(x-\xi)^{n-1}f^{(n-1)}(\xi)}{(n-1)!} \\ &= f(x) - \frac{(x-\xi)^n f^{(n)}[\xi + \theta(x-\xi)]}{n!}, \end{aligned} \tag{8*}$$

where ξ is a fixed point in $[-1, 1]$, and $0 < \theta < 1$. If x is in $[-1, 1]$, it follows from (5) and (8*) that

$$|P(x)| \leqq M_0 + 2^n \frac{M_n}{n!},$$

and hence by Lemma 2 that

$$|f^{(k)}(\xi)| = |P^{(k)}(\xi)| \leqq \frac{9kn^{2k}}{k!2^k}\left(M_0 + \frac{2^n M_n}{n!}\right) \tag{9*}$$

for $k = 1, 2, \cdots, n - 1$ and an arbitrary point ξ in $[-1, 1]$.

(II) Returning to (9), we distinguish two cases.

(IIa) We consider first the case in which

$$M_0 < \frac{M_n}{n!}, \qquad M_n' = M_n; \tag{10a}$$

see (6). If $0 < \lambda < 1$, the function $f(\lambda x)$ is defined for x in $[-1, 1]$, and we may apply (9) to $f(\lambda x)$ instead of to $f(x)$. We obtain

$$\lambda^k \,|\, f^{(k)}(0) \,| \leqq 3k^{\frac{1}{2}} n^k \left(M_0 + \lambda^n \frac{M_n}{n!}\right),$$

$$\begin{aligned} |\, f^{(k)}(0) \,| &\leqq 3k^{\frac{1}{2}} n^k \left(\frac{M_0}{\lambda^k} + \frac{\lambda^{n-k} e^n M_n}{n^n}\right) \\ &= 3k^{\frac{1}{2}}[M_0 t^{-k} + e^n M_n t^{n-k}]. \end{aligned} \tag{11}$$

We used the familiar inequality $n^n/n! < e^n$, and we defined t by

$$t = \frac{\lambda}{n},$$

so that we are free to choose t in the interval $(0, 1/n)$. We choose t so that the two terms in the square bracket in (11) become equal. This choice is admissible because, in virtue of (10a),

$$t = \frac{1}{e}\left(\frac{M_0}{M_n}\right)^{1/n} < \frac{1}{e}\left(\frac{1}{n!}\right)^{1/n} < \frac{1}{e}\,\frac{e}{n} = \frac{1}{n}.$$

We obtain from (11) by this choice

$$|\, f^{(k)}(0) \,| \leqq 3k^{\frac{1}{2}} 2M_0 e^k \left(\frac{M_n}{M_0}\right)^{k/n},$$

so that in case (10a) holds we have proved (7).

(IIb) We consider now the remaining case, in which

$$M_0 \geqq \frac{M_n}{n!}, \qquad M_n' = n!\,M_0. \tag{10b}$$

In this case, (9) yields directly

$$\begin{aligned} |\, f^{(k)}(0) \,| &\leqq 3k^{\frac{1}{2}} n^k 2M_0 \\ &= 6k^{\frac{1}{2}} n^k M_0^{1-k/n} \left(\frac{M_n'}{n!}\right)^{k/n} \\ &\leqq 6k^{\frac{1}{2}} n^k M_0^{1-k/n} \left(\frac{e}{n}\right)^k M_n'^{k/n}, \end{aligned}$$

so that we have proved (7) also under the condition (10b), and therefore completely.

(II*) In order to derive (7*) from (9*), we distinguish two cases. The first case is characterized by the condition

$$M_n^* = M_n > n!M_0 2^{-n}.$$

In this case, we consider $f[\xi + \lambda(x - \xi)]$ instead of $f(x)$, the point ξ being fixed in $[-1, 1]$; we choose λ in $(0, 1)$ and proceed as under (IIa). The remaining case is analogous to (IIb).

We consider now a theorem of a different character. Let $\inf \varphi(x)$ denote the greatest lower bound and let $\sup \varphi(x)$ denote the least upper bound of any real-valued function $\varphi(x)$ in $[a, b]$. With this notation, we have the following lemma.[11]

LEMMA 4. *Let the real-valued function $f(x)$ possess an n-th derivative in the interval $[a, b]$. Then*

$$\inf | f^{(n)}(x) | \leqq \frac{n!}{2}\left(\frac{4}{b - a}\right)^n \sup |f(x)|.$$

Let us consider, following Chebysheff, the polynomial

$$T(x) = \cos (n \arc \cos x) = 2^{n-1}x^n + \cdots . \tag{12}$$

Let c denote the center of $[a, b]$, so that $a + b = 2c$, and put

$$P(x) = T\left\{\frac{2(x - c)}{b - a}\right\} = \frac{1}{2}\left(\frac{4}{b - a}\right)^n x^n + \cdots . \tag{13}$$

If Lemma 4 is not true, $f^{(n)}(x)$ never vanishes in $[a, b]$, and it may then be supposed, without loss of generality,[12] to be positive in $[a, b]$. Moreover, if Lemma 4 is not true, there exists an H such that for all x in $[a, b]$

$$f^{(n)}(x) > \frac{n!}{2}\left(\frac{4}{b - a}\right)^n H > \frac{n!}{2}\left(\frac{4}{b - a}\right)^n |f(x)|. \tag{A}$$

Now consider

$$\varphi(x) = HP(x) - f(x). \tag{14}$$

We know that $P(x)$, defined by (12) and (13), takes at $n + 1$ points of $[a, b]$ (arranged in decreasing order) alternately the values 1 and -1; in particular, $P(b) = 1$. It follows from the second inequality (A) that $\varphi(x)$ takes alternately positive and negative values at the $n + 1$ points we have just mentioned, so that it vanishes at least at n different points of (a, b); and, in particular,

$$\varphi(b) = H - f(b) > 0. \tag{15}$$

[11] S. Bernstein, op. cit. 3b, p. 10. Our proof is a little different from the original one; we avoid using Fourier's rule. For another proof, see J. Shohat, *A simple proof of a formula of Tchebycheff*, Tôhoku Mathematical Journal, vol. 36(1932–1933), pp. 230–235.

[12] If $f^{(n)}(x)$ is never zero, it cannot take both positive and negative values, by a well-known theorem of Darboux.

We say that $\varphi(x)$ cannot vanish at more than n points of (a, b); otherwise, by Rolle's theorem, $\varphi'(x)$ would vanish at n points, $\varphi''(x)$ at $n - 1$ points, and so on; finally $\varphi^{(n)}(x)$ would vanish once. But (see (14) and (13))

$$\varphi^{(n)}(x) = H\,\frac{n!}{2}\left(\frac{4}{b-a}\right)^n - f^{(n)}(x) < 0 \qquad (a \leqq x \leqq b) \tag{16}$$

by the first inequality (A). Thus $\varphi(x)$ has exactly n zeros in (a, b), $\varphi'(x)$ has exactly $n - 1$, etc. Moreover, the $n - 1$ zeros of $\varphi'(x)$ must separate the n zeros of $\varphi(x)$; therefore, between the last zero of $\varphi(x)$ and the point b, $\varphi(x)$ and $\varphi'(x)$ must keep the same sign (they obviously have the same sign in a right-hand neighborhood of this last zero). Thus, by (15), we have

$$\varphi'(b) > 0.$$

In the same way we have $\varphi''(b) > 0, \varphi'''(b) > 0, \cdots$. But here we arrive at a contradiction, because, by (16), $\varphi^{(n)}(b) < 0$. To avoid the contradiction, we must discard (A), and so Lemma 4 is proved.

We use Lemma 4 to prove the following lemma.[13]

LEMMA 5. *If the real-valued function $f(x)$ possesses an n-th derivative $f^{(n)}(x)$ which is monotonic in $[a, b]$, if $|f(x)| \leqq M$ in $[a, b]$, and if $0 < l < \frac{1}{2}(b - a)$, then, in $[a + l, b - l]$,*

$$|f^{(n)}(x)| \leqq \frac{n!}{2}\left(\frac{4}{l}\right)^n M. \tag{17}$$

Because $f^{(n)}(x)$ is monotonic, the maximum of $|f^{(n)}(x)|$ in $[a + l, b - l]$ is attained at one of the end-points of that interval, and this maximum coincides with the minimum of $|f^{(n)}(x)|$ in the closed interval of length l which is separated from $[a + l, b - l]$ by the end-point in question.[14] Applying Lemma 4 to this interval of length l, we obtain (17).

It may contribute to our understanding of Lemma 5 to compare it with the following lemma on analytic functions.

If the analytic function $f(z)$ is regular, $|f(z)| \leqq M$ in a circle of radius r, and $0 < l < \frac{1}{2}r$, then in the concentric circle of radius $r - l$

$$|f^{(n)}(z)| \leqq n!Ml^{-n}.$$

4. **A new lemma on derivatives of real functions.** We are now prepared to prove

[13] See E. Landau, *Über einen Satz von Herrn Esclangon*, Mathematische Annalen, vol. 102(1929–1930), pp. 177–188.

[14] $f^{(n)}(x)$, as a monotonic derivative, is continuous by the theorem of Darboux quoted in footnote 12.

LEMMA 6. *If p and q are positive integers, and $g(x)$ is real-valued, possesses a $(p + 2q)$-th derivative, and satisfies*

$$|g(x)| \leqq M, \qquad g^{(p+2q)}(x) \leqq 0 \tag{18}$$

in $[-1, 1]$, then

$$g^{(p)}(x) \leqq A^{p+2q}(p + 2q)^p M$$

in $[-1, 1]$, where $A = 30e^{30e+1}$.

The main point is that A is an absolute constant, independent of the choice of $g(x)$, p, and q. It is also important that the estimate for $g^{(p)}(x)$ holds throughout $[-1, 1]$, and not merely in a sub-interval.

We put

$$p + 2q - 1 = n. \tag{19}$$

By the second condition (18), $g^{(n)}(x)$ is steadily decreasing. Therefore we may apply Lemma 5; we obtain from (17)

$$|g^{(n)}(x)| < n!\left(\frac{4}{h}\right)^n M \qquad (-1 + h \leqq x \leqq 1 - h),$$

where $0 < h < 1$. By this inequality and the first condition (18), the hypothesis (5) of Lemma 3 is fulfilled in $[-1 + h, 1 - h]$. In order to simplify the application of Lemma 3, we choose h, taking (6) into consideration, so that

$$n!\left(\frac{4}{h}\right)^n M = n!M(1 - h)^{-n};$$

i.e., we put $h = 4/5$. With this choice (7) yields

$$\begin{aligned} |g^{(k)}(0)| &\leqq 6k^{\frac{1}{2}}e^k M^{1-k/n}(n!M5^n)^{k/n}, \\ |g^{(k)}(0)| &\leqq (30en)^k M \end{aligned} \tag{20}$$

for $k = 1, 2, 3, \cdots, n$.

By Taylor's formula

$$g^{(p)}(x) = g^{(p)}(0) + \frac{xg^{(p+1)}(0)}{1!} + \cdots + \frac{x^{2q-1}g^{(p+2q-1)}(0)}{(2q-1)!} + \frac{x^{2q}g^{(p+2q)}(\xi)}{(2q)!},$$

where $0 < \xi < x$. The last term is not positive, by the second condition (18). Using (20) for the other terms (see (19)), we obtain in $[-1, 1]$

$$\begin{aligned} g^{(p)}(x) &\leqq (30en)^p M\left(1 + \frac{30en}{1!} + \frac{(30en)^2}{2!} + \cdots\right) \\ &= (30en)^p M e^{30en} \\ &\leqq (30e \cdot e^{30e})^n n^p M. \end{aligned}$$

This, with (19), proves Lemma 6.

5. **Proof of Theorem** 1. We consider the function $f(x)$ of Theorem 1. Let M_n denote the maximum of $|f^{(n)}(x)|$ in $[-1, 1]$.

We know that both $f^{(n_k)}(x)$ and $f^{(n_k+2q_k)}(x)$ keep a constant sign in $[-1, 1]$. We may assume, without loss of generality, that

$$(21) \qquad f^{(n_k+2q_k)}(x) \leqq 0$$

in $[-1, 1]$. It follows that

$$(22) \qquad f^{(n_k)}(x) \geqq 0$$

in $[-1, 1]$. We now apply Lemma 6, putting

$$g(x) = f^{(n_{k-1})}(x), \qquad p = n_k - n_{k-1}, \qquad q = q_k, \qquad M = M_{n_{k-1}}.$$

Condition (18) is satisfied since we have (21). Using (22) also, we obtain from Lemma 6 that

$$(23) \qquad M_{n_k} \leqq A^{n_k-n_{k-1}+2q_k}(n_k - n_{k-1} + 2q_k)^{n_k-n_{k-1}}M_{n_{k-1}}.$$

We also have

$$(24) \qquad M_{n_1} \leqq A^{n_1+2q_1}(n_1 + 2q_1)^{n_1}M,$$

M being the maximum of $|f(x)|$ in $[-1, 1]$. We may consider (24) as the special case of (23) in which $k = 0$, if we write $n_0 = 0$. By (24) and repeated application of (23) we obtain

$$(25) \qquad M_{n_k} \leqq MA^{n_k+2(q_1+q_2+\cdots+q_k)} \prod_{p=1}^{k} (n_p - n_{p-1} + 2q_p)^{n_p-n_{p-1}}.$$

This inequality (25) supplies us with an estimate for the derivatives $f^{(n_1)}(x)$, $f^{(n_2)}(x), \cdots, f^{(n_k)}(x), \cdots$ in $[-1, 1]$. Starting from this, we can estimate the remaining derivatives at $x = 0$, using (7) of Lemma 3. This is the main idea of the proof; in order to supply the details, we must consider the cases (I), (II), and (III) separately.

(I) The particular hypothesis characterizing case (I) is the existence of a positive constant B such that, for $k = 1, 2, 3, \cdots$,

$$(26) \qquad n_k - n_{k-1} < B, \qquad q_k < B.$$

From this and (25), we obtain

$$(27) \qquad M_{n_k} \leqq M(A^{1+2B}3B)^{n_k} = MC^{n_k}.$$

We now apply inequality (7) of Lemma 3 to $f^{(n_k)}(x)$ instead of $f(x)$, in the interval $[-1, 1]$, with $n_{k+1} - n_k$ and h $(h = 1, 2, \cdots, n_{k+1} - n_k)$ instead of n and k. We must first (see (6)) estimate

$$\max(M_{n_{k+1}}, (n_{k+1} - n_k)!M_{n_k}).$$

We find, using (26) and (27),

$$\max\left(M_{n_{k+1}},\ (n_{k+1}-n_k)!M_{n_k}\right) \leqq \max\left(MC^{n_{k+1}},\ B^{n_{k+1}-n_k}MC^{n_k}\right) = MC^{n_{k+1}},$$

since, by (27), $B < C$.

Lemma 3 now yields

$$|f^{(n_k+h)}(0)| \leqq 6h^{\frac{1}{2}}e^h M_{n_k}^{1-h/(n_{k+1}-n_k)}\left(MC^{n_{k+1}}\right)^{h/(n_{k+1}-n_k)}. \tag{28}$$

Using (26) and (27) again, we obtain

$$|f^{(n_k+h)}(0)| \leqq 6B^{\frac{1}{2}}e^B MC^{n_k+h} \qquad (h = 1, 2, \cdots, n_{k+1}-n_k);$$

so that we have, with an appropriate constant D,

$$|f^{(n)}(0)| < DC^n \qquad (n = 0, 1, 2, \cdots). \tag{29}$$

By (29), we see that the Maclaurin series of $f(x)$ converges everywhere, and is the Maclaurin series of an entire function of exponential type. But this series actually represents $f(x)$ in the interval $[-1, 1]$ because the remainder after the term containing x^{n_k-1} approaches zero (use (27) and Lagrange's form of the remainder).

Thus we have proved case (I) of Theorem 1. Moreover, we have obtained a pattern which we may follow in proving the remaining cases. First, we estimate the derivatives $f^{(n_1)}(x), f^{(n_2)}(x), \cdots, f^{(n_k)}(x), \cdots$ in the whole interval $[-1, 1]$, using (25). Second, we estimate the other derivatives at the point 0, using inequality (7) of Lemma 3 as we used it here for (28). These estimates must show that the coefficients of the Maclaurin series of $f(x)$ have orders of magnitude according with the respective statements of Theorem 1. The argument showing that the series actually represents $f(x)$ in $[-1, 1]$ is the same in all cases, and need not be repeated.

(II) The particular hypothesis characterizing case (II) is the existence of a positive constant B such that, for $k = 1, 2, 3, \cdots$,

$$n_k - n_{k-1} < Bn_k^{\delta}, \qquad q_k < Bn_k^{\delta}, \tag{30}$$

$$q_1 + q_2 + \cdots + q_k < Bn_k. \tag{31}$$

Then (25) yields

$$\begin{aligned} M_{n_k} &\leqq MA^{n_k+2Bn_k}\prod_{p=1}^{k}(3Bn_k^{\delta})^{n_p-n_{p-1}} \\ &= M(A^{1+2B}3B)^{n_k}n_k^{\delta n_k} \\ &= MC^{n_k}n_k^{\delta n_k}. \end{aligned} \tag{32}$$

In order to prepare for the application of Lemma 3, we consider

$$\max\left(M_{n_{k+1}},\ (n_{k+1}-n_k)!M_{n_k}\right) \leqq \max\left(MC^{n_{k+1}}n_{k+1}^{\delta n_{k+1}},\ (Bn_{k+1}^{\delta})^{n_{k+1}-n_k}MC^{n_k}n_k^{\delta n_k}\right) = MC^{n_{k+1}}n_{k+1}^{\delta n_{k+1}}.$$

Here we have used (30) and (32); by the latter, $B < C$.

We now apply inequality (7) of Lemma 3 to

$$f^{(n_k)}(x), \qquad [-1, 1], \qquad n_{k+1} - n_k, \qquad h$$

instead of to

$$f(x), \qquad [-l, l], \qquad n, \qquad k;$$

we obtain, using (32),

$$\begin{aligned} |f^{(n_k+h)}(0)| &\leqq 6h^{\frac{1}{2}} e^h M_{n_k}^{1-h/(n_{k+1}-n_k)} (MC^{n_{k+1}} n_{k+1}^{\delta n_{k+1}})^{h/(n_{k+1}-n_k)} \\ &\leqq (6e)^h MC^{n_k+h} (n_k + h)^{\delta(n_k+h)} e^{\delta R}, \end{aligned} \tag{33}$$

where

$$R = (1 - \alpha)\varphi(n) + \alpha\varphi(N) - \varphi(\nu); \tag{34}$$

we are using the abbreviations

$$n_k = n, \qquad n_{k+1} = N, \qquad h/(n_{k+1} - n_k) = \alpha, \qquad n_k + h = \nu, \tag{35}$$

$$x \log x = \varphi(x), \tag{36}$$

so that

$$(1 - \alpha)n + \alpha N = \nu. \tag{37}$$

But,

$$\begin{aligned} \varphi(n) &= \varphi(\nu) + (n - \nu)\varphi'(\nu) + \tfrac{1}{2}(n - \nu)^2\varphi''(\nu_1), \\ \varphi(N) &= \varphi(\nu) + (N - \nu)\varphi'(\nu) + \tfrac{1}{2}(N - \nu)^2\varphi''(\nu_2), \end{aligned}$$

with $n < \nu_1 < \nu < \nu_2 < N$; and therefore since (see (36)) $\varphi''(x) = 1/x$ is a decreasing function, it follows from (34) and (37) that

$$\begin{aligned} R &= \tfrac{1}{2}[(1 - \alpha)(n - \nu)^2\varphi''(\nu_1) + \alpha(N - \nu)^2\varphi''(\nu_2)] \\ &< \tfrac{1}{2}\alpha(1 - \alpha)(N - n)^2\varphi''(n) \leqq \tfrac{1}{8}(N - n)^2\varphi''(n). \end{aligned}$$

Returning to (36), (35), (30), we obtain

$$R < (n_{k+1} - n_k)^2 n_k^{-1} < B^2 n_{k+1}^{2\delta} n_k^{-1}, \qquad n_{k+1} < n_k(1 - Bn_{k+1}^{\delta-1})^{-1}.$$

Using this, we see from (33) and (32) that there is a constant D such that for $n = 1, 2, 3, \cdots$

$$|f^{(n)}(0)| < D^n n^{\delta n}.$$

By this inequality, the Maclaurin series of $f(x)$ is that of an entire function of finite order not exceeding $1/(1 - \delta)$.

(III) This case is characterized by the conditions

$$\lim_{k\to\infty} \frac{n_{k+1}}{n_k} = 1, \qquad \lim_{k\to\infty} \frac{q_k}{n_k} = 0,$$

and (31). The application of (25) and of Lemma 3 to this case follows closely the developments in case (II); carrying through the calculation, we are led

again to the consideration of the same expression (34). So we leave this case to the reader, who may show that, given any positive ϵ, we have for all sufficiently large k and n

$$M_{n_k} < (\epsilon n_k)^{n_k}, \qquad |f^{(n)}(0)| < (\epsilon n)^n.$$

6. **Examples and comments.** Before discussing the separate cases of Theorems 1 and 2, we shall make a few rather obvious remarks concerning all cases.

Since all properties of $f(x)$ considered in Theorems 1 and 2 are unchanged by a non-fractional linear transformation, these theorems do not change substantially when we replace $[-1, 1]$ by any other closed interval. Any function $f(x)$ which satisfies the hypothesis of any case of Theorem 2 also satisfies the hypothesis of the corresponding case of Theorem 1, if we put $\frac{1}{2}(n_{k+1} - n_k) = q_k$. Consequently we are at liberty to choose another interval instead of $[-1, 1]$, or to content ourselves with the consideration of one of Theorems 1 and 2 in the following discussion.

(I) All cases of Theorem 2 are concerned with a function $f(x)$ having an infinite sequence of derivatives of even order, $f^{(n_1)}(x), f^{(n_2)}(x), \cdots$, which do not change sign in a fixed interval, the signs being alternately $+$ and $-$. The case (I) adds the particular hypothesis that the sequence $n_1, n_2, \cdots$ does not increase more rapidly than an arithmetic progression, in the sense that $n_k - n_{k-1}$ remains bounded; and draws the particular conclusion that the growth of $f(x)$ does not exceed some finite type of order one.

Can we draw the same conclusion from a less restricted hypothesis? We have no complete answer to this question. However, the example which we shall discuss under (II) of this section shows that if the difference $n_k - n_{k-1}$ increases as slowly as k^ϵ, where ϵ is a fixed positive number, $f(x)$ may have order greater than one.

Can we draw a stronger conclusion from the same hypothesis? Here we have a complete answer. No hypothesis whatever on the signs of the derivatives in a fixed interval can imply that the growth of the function is less than order one and finite type. In fact, given any sequence $\epsilon_0, \epsilon_1, \cdots, \epsilon_n, \cdots$, where $\epsilon_n = 1$ or -1, there exists a function $g(x)$ of order and type 1, such that $g^{(n)}(x)$ has the sign of ϵ_n in $(-\log 2, \log 2)$ for $n = 0, 1, 2, \cdots$. Such a function $g(x)$ is defined by

$$g(x) = \epsilon_0 + \frac{\epsilon_1 x}{1!} + \frac{\epsilon_2 x^2}{2!} + \cdots + \frac{\epsilon_n x^n}{n!} + \cdots. \tag{38}$$

For,

$$g^{(n)}(x) = \epsilon_n + \frac{\epsilon_{n+1} x}{1!} + \frac{\epsilon_{n+2} x^2}{2!} + \cdots,$$

$$\epsilon_n g^{(n)}(x) \geqq 1 - \frac{|x|}{1!} - \frac{|x|^2}{2!} - \cdots = 2 - e^{|x|} > 0 \tag{39}$$

for $|x| < \log 2$.

Consideration of the particular function (38) leads to a complement which applies to all three cases of our theorems.[15]

Theorems 1 *and* 2 *remain valid if we change the inequalities which the derivatives of* $f(x)$ *are assumed to satisfy in the following way.*

Instead of $f^{(n)}(x) \geqq 0$ *we assume only* $f^{(n)}(x) \geqq -p^n$, *instead of* $f^{(n)}(x) \leqq 0$ *we assume only* $f^{(n)}(x) \leqq p^n$, *where* p *is a fixed positive number.*

First, by considering $f(x/p)$ instead of $f(x)$, we can reduce the general case to the special case where $p = 1$. Second, it is sufficient to prove the theorem in an interval $[-l, l]$, where $l \leqq \frac{1}{2}$, because shifting the interval does not change anything that matters, and any ("large") interval can be covered by a finite number of ("slightly") overlapping intervals of length not exceeding 1. Now construct the sequence $\epsilon_0, \epsilon_1, \cdots, \epsilon_n, \cdots$ according to the following rule.

If it is assumed that $f^{(n)}(x) \geqq -1$, take $\epsilon_n = 1$; if it is assumed that $f^{(n)}(x) \leqq 1$, take $\epsilon_n = -1$; and take ϵ_n arbitrarily, e.g., $\epsilon_n = 1$, if nothing is assumed about $f^{(n)}(x)$. With these numbers ϵ_n we define $g(x)$ by (38). Finally, we consider

$$h(x) = f(x) + g(x)/(2 - e^{\frac{1}{2}}).$$

Inequality (39) shows that $h^{(n)}(x) \geqq 0$ or $h^{(n)}(x) \leqq 0$ in $[-l, l]$ according as it was originally assumed that $f^{(n)}(x) \geqq -1$ or $f^{(n)}(x) \leqq 1$. Then the original form of Theorem 1 (or Theorem 2) may be applied to $h(lx)$.

(II) *There exist entire functions of order exactly* $1/(1-\delta)$ *satisfying the hypothesis of case* (II) *of Theorem* 2 (*and, therefore, also the hypothesis of case* (II) *of Theorem* 1).

In other words, the conclusion concerning the order of $f(x)$ which we deduced from the hypothesis of case (II) is the strongest possible.

We are given δ, $0 < \delta < 1$. We put

$$\frac{1}{1-\delta} = \rho, \tag{40}$$

$$2[k^\rho] = n_k \qquad (k = 1, 2, 3, \cdots); \tag{41}$$

we use from now on the usual convention that $[r]$ denotes the integral part of the real number r. It is evident that the function

$$f(x) = \sum_{p=1}^{\infty} (-1)^p \frac{n_p^{\delta n_p} x^{n_p}}{n_p!} \tag{42}$$

is entire and that its order is exactly $1/(1 - \delta) = \rho$. We shall show that it satisfies the hypothesis of case (II) of Theorem 2.

The integers $n_1, n_2, \cdots$ defined by (41) are even, and (see (40))

$$\lim_{k\to\infty} (n_k - n_{k-1})n_k^{-\delta} = \rho 2^{1-\delta}.$$

[15] This complement was found in a conversation between one of the authors and Professor E. Artin.

Thus the condition $n_k - n_{k-1} = O(n_k^\delta)$ is certainly satisfied. We obtain from (42) that

$$\frac{(-1)^k f^{(n_k)}(x)}{n_k^{\delta n_k}} = 1 + \sum_{\nu=1}^{\infty} (-1)^\nu \frac{n_{k+\nu}^{\delta n_{k+\nu}}}{n_k^{\delta n_k}} \frac{x^{n_{k+\nu}-n_k}}{(n_{k+\nu} - n_k)!}. \tag{43}$$

We introduce the abbreviations

$$n_k = n, \qquad n_{k+\nu} = N, \tag{44}$$

and estimate the absolute value of the general term of the series thus:

$$\begin{aligned} \left(\frac{N^N}{n^n}\right)^\delta \frac{|x|^{N-n}}{(N-n)!} &< \left(1 + \frac{N-n}{n}\right)^{n\delta} \frac{N^{(N-n)\delta} |x|^{N-n}}{\{2\pi(N-n)\}^{\frac{1}{2}}(N-n)^{N-n} e^{-(N-n)}} \\ &< \frac{\{e^{1+\delta} |x|\}^{N-n}}{\{2\pi(N-n)\}^{\frac{1}{2}}} \left(\frac{N^\delta}{N-n}\right)^{N-n}. \end{aligned} \tag{45}$$

Now $x^\delta/(x - n)$ is a decreasing function of x for $x > n$, as is easily shown by differentiation. Therefore, by (44) and (41),

$$\begin{aligned} \frac{N^\delta}{N-n} \leqq \frac{n_{k+1}^\delta}{n_{k+1} - n_k} &< \frac{2^\delta (k+1)^{\rho\delta}}{2[(k+1)^\rho - 1 - k^\rho]} \\ &< \frac{2^{\delta-1}(k+1)^{\rho\delta}}{\rho k^{\rho-1} - 1} < 1 \end{aligned} \tag{46}$$

for sufficiently large k; observe that, by (40), $\rho\delta = \rho - 1$. From this inequality and (45) it follows that, for large k, (43) is certainly positive in $[-\frac{1}{2}e^{-1-\delta}, \frac{1}{2}e^{-1-\delta}]$. This shows that the function (42) satisfies all the requirements of case (II).

(III) The hypothesis of case (III) of Theorem 2 concerning the sequence $n_1, n_2, \cdots, n_k, \cdots$, namely that

$$\lim_{k\to\infty} \frac{n_{k+1}}{n_k} = 1,$$

cannot be replaced by the weaker hypothesis

$$\limsup_{k\to\infty} \frac{n_{k+1}}{n_k} = a,$$

where $a > 1$, without invalidating the conclusion that $f(x)$ is an entire function. In fact the following statement is true.

If $a > 1$, there exist an increasing sequence of even numbers $n_1, n_2, \cdots, n_k, \cdots$ and a function $f(x)$, analytic in a given interval but not entire, such that $(-1)^k f^{(n_k)}(x) > 0$ in that interval and

$$\lim_{k\to\infty} \frac{n_{k+1}}{n_k} = a. \tag{47}$$

The construction is much the same as that of the preceding example. This time we put

$$n_k = 2[a^k],$$

and the function $f(x)$ is again given by (42), but with δ replaced by 1. Then evidently the radius of convergence of the series is finite (in fact, $1/e$). Following (44), (45), (46) with obvious changes, we show that, for sufficiently large k, (43) is certainly positive in $[-(a-1)/(2ae^2), (a-1)/(2ae^2)]$.

(IV) One might ask whether a function satisfying the hypothesis of Theorem 2 necessarily has all its derivatives of constant sign in *some* interval, not necessarily the whole of $[-1, 1]$. If this were the case, Theorem 2 would be contained in Bernstein's results. However, the following example shows that *for any increasing sequence of integers $n_1, n_2, \cdots$ with the property $n_{k+1} - n_k \geqq 2$ and for any sequence $\epsilon_1, \epsilon_2, \cdots$ ($\epsilon_k = \pm 1$), there is a function $f(x)$ such that $\epsilon_k f^{(n_k)}(x) > 0$ in $[-1, 1]$, while the points at which the derivatives $f^{(n_k-1)}(x)$ change sign form a set which is everywhere dense in $[-1, 1]$.*

Let $a_0, a_1, a_2, \cdots$ be points of $[-1, 1]$. Define $P_0(x) = 1$,

$$P_n(x) = \int_{a_0}^{x} dx_1 \int_{a_1}^{x_1} dx_2 \cdots \int_{a_{n-1}}^{x_{n-1}} dx_n \qquad (n = 1, 2, \cdots).$$

Then evidently

$$(48) \qquad | P_n(x) | \leqq 2^n \qquad (-1 \leqq x \leqq 1; n = 0, 1, 2, \cdots),$$

$$(49) \qquad P_n^{(n)}(x) = 1.$$

Now let $a_{n_1-1}, a_{n_2-1}, \cdots$ be a sequence of numbers everywhere dense in $[-1, 1]$; let $a_n = 0$ when n is not a member of the sequence $n_1 - 1, n_2 - 1, \cdots$. Let $A_{n_k} = \epsilon_k b^{n_k}$ ($k = 1, 2, \cdots$), with $0 < b < \frac{1}{4}$; let $A_n = 0$ when n is not a member of the sequence $n_1, n_2, \cdots$. Then

$$(50) \qquad f(x) = \sum_{n=0}^{\infty} A_n P_n(x)$$

has the desired property. For, since for $k = 1, 2, \cdots, n-1$,

$$P_n^{(k)}(x) = \int_{a_k}^{x} dx_1 \int_{a_{k+1}}^{x_1} dx_2 \cdots \int_{a_{n-1}}^{x_{n-k-1}} dx_{n-k},$$

by (48) the series in (50) is dominated by $\sum_{n=0}^{\infty} (2b)^n$, and the formal series $f^{(k)}(x)$ is dominated by $\sum_{n=k}^{\infty} b^n 2^{n-k}$. Hence $f^{(k)}(x)$ may be obtained by termwise differentiation of (50). For x in $[-1, 1]$, by (49),

$$f^{(n_k)}(x) = A_{n_k} + \sum_{n=n_k+1}^{\infty} A_n P_n^{(n_k)}(x);$$

by (50),

$$\epsilon_k f^{(n_k)}(x) \geqq b^{n_k} - \sum_{n=n_k+1}^{\infty} b^n 2^{n-n_k}$$

$$= b^{n_k}\left\{1 - \frac{2b}{1-2b}\right\} > 0.$$

On the other hand, $\epsilon_k f^{(n_k-1)}(x)$ increases in $[-1, 1]$, has the value $A_{n_k-1} = 0$ at $x = a_{n_k-1}$, and hence changes sign at a_{n_k-1}.

(V) If in cases (I) and (II) of Theorems 1 and 2 we require only that the inequalities involving $f^{(n_k)}(x)$ hold in an interval $[-l_k, l_k]$, where l_k does not approach zero too rapidly, we can still show that the Maclaurin series of $f(x)$ is the Maclaurin series of an entire function. Sufficient restrictions on l_k in the two cases are

(I) $$kl_k \to \infty \qquad (k \to \infty),$$

(II) $$n_k^{1-\delta} l_k \to \infty \qquad (k \to \infty).$$

Hence we may state, for example,

If $f(x)$ is analytic[16] *in $[-1, 1]$, if in $[-l_k, l_k]$ $f^{(n_k)}(x)$ and $f^{(n_k+2q_k)}(x)$ do not change sign and $f^{(n_k)}(x)f^{(n_k+2q_k)}(x) \leqq 0$, where $n_k - n_{k-1} = O(1)$, $q_k = O(1)$, and $kl_k \to \infty$, then $f(x)$ is an entire function.*

It will be clear that the order of the entire function will be smaller, the less rapidly $l_k \to 0$.

To establish our assertion, we have only to make slight modifications in the proof of Theorem 1. It is convenient to suppose that $l_1 > l_2 > \cdots$. We apply Lemma 6 to the function $g(x) = f^{(n_k-1)}(l_k x)$. Inequality (23) becomes

$$M_{n_k} \leqq l_k^{-(n_k-n_{k-1})} A^{n_k-n_{k-1}+2q_k}(n_k - n_{k-1} + 2q_k)^{n_k-n_{k-1}} M_{n_{k-1}},$$

where M_{n_p} now denotes the maximum of $|f^{(n_p)}(x)|$ in $[-l_p, l_p]$. Since l_k is a decreasing sequence, (25) is replaced by

$$M_{n_k} \leqq M l_k^{-n_k} A^{n_k+2(q_1+\cdots+q_k)} \prod_{p=1}^{k} (n_p - n_{p-1} + 2q_p)^{n_p-n_{p-1}}.$$

Using (26), we obtain

$$M_{n_k} \leqq M\left(\frac{C}{l_k}\right)^{n_k}.$$

Following the rest of the proof of case (I) of Theorem 1, we find that for $h = 0, 1, \cdots, n_{k+1} - n_k$,

$$|f^{(n_k+h)}(0)| \leqq E\left(\frac{C}{l_{k+1}}\right)^{n_k+h}$$

[16] Or even belongs to a Denjoy-Carleman quasi-analytic class.

with a suitable constant E. Hence

$$\frac{1}{n_k + h} \,|\, f^{(n_k+h)}(0) \,|^{1/(n_k+h)} \to 0$$

if

$$\frac{C}{l_{k+1}\, n_k} \to 0,$$

i.e., if

$$n_k l_{k+1} \to \infty.$$

But since $k \leqq n_k \leqq kB$, this is equivalent to

$$k l_k \to \infty.$$

Case (II) is treated similarly.

DUKE UNIVERSITY AND BROWN UNIVERSITY.

THE DISTRIBUTION OF PRIMES

By Aurel Wintner

1. **Simple prime factors.** If $f(n)$ is a function of the positive integer n, let f_t denote the set of the solutions n of $f(n) = t$, and let $f_t(x)$ be the number of those elements of f_t which are less than x.

Thus, if $f(n)$ is the number of distinct primes dividing n or is 0 according as n is or is not square-free, then n is in f_0 if and only if it is not square-free so that $f_0(x) \sim (1 - \zeta(2)^{-1})x$ as $x \to \infty$. On the other hand, if $m > 0$, then $f_m(x)$ is the number of those integers n less than x which are composed of exactly m distinct prime factors, a number usually denoted by $\pi_m(x)$. Apparently, it was observed already by Gauss[1] that the prime number theorem, i.e., $\pi_1(x) \sim x(\log x)^{-1}$, implies, for every fixed m $(= 1, 2, \cdots)$, the asymptotic relation

$$\pi_m(x) \sim L_m(x), \tag{1}$$

where

$$L_m(x) = \frac{x(\log x)^{-1}(\log \log x)^{m-1}}{(m-1)!}.$$

Thus $L_1(x) + L_2(x) + \cdots \equiv x$, although

$$\pi_1(x) + \pi_2(x) + \cdots \equiv [x] - f_0(x) \sim x/\zeta(2).$$

The latter anomaly presents itself also in case of the function $f(n) = \theta(n)$ which plays a central rôle in the following considerations and represents the number of *simple* prime factors of n (for instance, $\theta(15) = 2$, $\theta(60) = 2$, $\theta(24) = 1$). Clearly, there exists for every n exactly one m for which the set θ_m contains n so that $\theta_1(x) + \theta_2(x) + \cdots \equiv [x] \sim x$. However, for every fixed m,

$$\theta_m(x) \sim \text{const. } L_m(x), \tag{2}$$

where

$$\text{const.} = \frac{\zeta(2)\zeta(3)}{\zeta(6)}.$$

In fact, if m is fixed, an n is in θ_m if and only if $n = p_1 \cdots p_m j$ holds for m distinct primes $p_1, \cdots, p_m$ and for a j having only multiple prime factors each of which is distinct from $p_1, \cdots, p_m$. Since $\pi_m(x)$ is the number of those integers less than x which are of the form $p_1 \cdots p_m$, it follows that, in order to pass from (1) to (2), it is sufficient to show that $\sum 1/i$ has a finite

Received January 7, 1942.

[1] C. F. Gauss, *Werke*, vol. 10, part 1, 1917, p. 11 and p. 17. For the remainder term, cf. E. Landau, *Über die Verteilung der Zahlen, welche aus ν Primfaktoren zusammengesetzt sind*, Göttingen Nachrichten, 1911, pp. 361–381.

value, $\zeta(2)\,\zeta(3)/\zeta(6)$, where i runs through the set of *all* positive integers possessing no simple prime factors. But the product of two arbitrary elements of the latter set is an element of this set. Furthermore, a prime power p^k is in the set if and only if $k \neq 1$. Hence, by Euler's factorization, the sum $\sum 1/i$ is the product of all factors $1 + 0 + p^{-2} + p^{-3} + \cdots$. Since the latter sum, being equal to $1 + p^{-2}/(1 - p^{-1})$, is identical with the reciprocal value of $(1 - p^{-2})(1 - p^{-3})/(1 - p^{-6})$, the assertion follows by applying to $s = 2, 3, 6$ the product definition of $\zeta(s)$.

These remarks will now be combined with certain general facts regarding additive functions.

2. **Additive functions.** A function $f(n)$ is called additive if

$$f(n_1 n_2) = f(n_1) + f(n_2) \tag{3}$$

whenever

$$(n_1, n_2) = 1,$$

i.e., n_1 and n_2 are relatively prime (in particular $f(1) = 0$). Thus an additive $f(n)$ can uniquely be characterized by an arbitrary assignment of the double sequence formed by the values $f(q^k)$, where q and k run through all primes and through all positive integers respectively. It also is clear that, if $f^p(n)$ denotes, for a fixed prime p, that additive function for which the value $f^p(q^k)$ is 0 or $f(p^k)$ according as the prime q is or is not distinct from p, then $f(n) = \sum f^p(n)$ for every n (it being understood that, although the summation runs through all primes p, the sum is finite for every n, since $f^p(n) = 0$ whenever p exceeds a bound depending on n). It is known[2] that, if $f(n)$ is any additive function, the functions $f^p(n), f^r(n), \cdots$ belonging to any finite set of distinct primes $p, r, \cdots$ are statistically independent and that

$$\int_{-\infty}^{\infty} \exp(i\alpha u)\, d\phi^p(\alpha) = (1 - p^{-1}) \sum_{k=0}^{\infty} p^{-k} \exp(if(p^k)u) \quad (-\infty < u < \infty), \tag{4}$$

where $\phi^p = \phi^p(\alpha)$ $(-\infty < \alpha < \infty)$ is the asymptotic distribution function of the additive function $f^p(n)$ of n.

3. **Poisson's law.** Since the terms of the series $f(n) = \sum f^p(n)$ are *statistically independent*, and since the appearance of primes for which $f^p(n)$ does not vanish is a *rare* event when $n \to \infty$, it is to be expected that, if the functions $f^p(n)$ belonging to the various primes p are sufficiently *homogeneous* with respect to p (in particular, if $f^p(p) = 1$ for every p), then the structure of the additive function $f(n)$ is subject to conditions very similar to those under which Poisson's law of distribution can rigorously be deduced.[3]

[2] Cf. P. Erdös and A. Wintner, *Additive arithmetical functions and statistical independence*, American Journal of Mathematics, vol. 61 (1939), pp. 713–721, more particularly, pp. 718–719.

[3] Cf., e.g., A. Khintchine, *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 2 (1933), no. 4, chap. 2, pp. 16–20.

A "random variable" is said to be distributed according to Poisson's law if the distinct "states" of which it is capable form an infinite sequence $S_1, S_2, \cdots$ in such a way that the "probability", say $[S_m]$, of the "state" S_m has the value $[S_m] = e^{-\lambda}\lambda^{m-1}/(m-1)!$, where λ is a positive number independent of m $(= 1, 2, \cdots)$. Clearly, the total probability, $\sum [S_m]$, is 1 for every λ. Since, as easily verified, $\sum m^2[S_m] - (\sum m[S_m])^2 = \lambda$, the square root of λ is precisely the standard deviation.

Now let the "random variable" be identified with an additive function $f(n)$ and let its m-th "state" S_m take place for those values of n for which $f(n)$ attains the value m. Thus $f(n)$ is one of the "states" $S_1, S_2, \cdots$ for every $n > 1$ if and only if every $f(q^k)$ is a positive integer. Actually, the values attained for prime powers q^k in which $k > 1$ are relatively unimportant. In fact, all that is needed is that the "probability" that $f(n)$ be in none of the "states" S_m be 0; in other words, that only $o(x)$ of the first x positive integers n should lead to values $f(n)$ distinct from every positive integer m. It is clear from the definition of $f_t(x)$ at the beginning of this paper that the ratio $f_m(x)/x$ is the relative frequency ("probability a posteriori") of the "state" S_m, when n varies from $n = 1$ to $n = x$ ($n = 1$ and $n = x$ being excluded).

4. **The problem.** If n could be restricted to the range $1 < n < x$, one might infer that $f_m(x)/x$ is identical with the value $[S_m]$ ("probability a posteriori"), supplied by the deduction of Poisson's law. But if n is restricted to a finite range $1 < n < x$, then the terms of $f(n) = \sum f^p(n)$ are not statistically independent, since their statistical independence holds only in terms of the asymptotic distributions belonging to $x \to \infty$. Naturally, one can consider the infinite n-range, on which the functions $f^{(p)}(n)$ are statistically independent in this asymptotic sense $(x \to \infty)$. But then Poisson's law is not available since its deduction assumes finite Lebesgue measures (which are additive) and not *asymptotic* distributions (defining relative measures which are not, in general, additive). Accordingly, it cannot be inferred that, if m is fixed and $x \to \infty$, then the relative frequency of the "state" S_m is asymptotically represented by the corresponding Poissonian probability, i.e., that

$$\frac{f_m(x)}{x} \sim \frac{e^{-\lambda}\lambda^{m-1}}{(m-1)!} \tag{5}$$

as $x \to \infty$, where $\lambda = \lambda(x)$; it is understood that $\lambda = \lambda(x)$ denotes the square of the standard deviation belonging to the finite range preceding x. Nevertheless, the remarks made after (3) suggest that (5) is likely to be true for every fixed m if the underlying additive function $f(n)$ satisfies certain Tauberian conditions; conditions which make legitimate the interchange of two limit processes (the latter are the step $x \to \infty$ and the process occurring in the standard deduction of Poisson's law). It is clear from the purely Tauberian character of the problem that no argument based on probability can prove the heuristic relation (4). Actually, it turns out that, even in case of the simplest functions $f(n)$, the validity of Poisson's law (4) is equivalent to the prime number theorem,

$\pi(x) \sim x/\log x$. The conditions for the additive function $f(n)$ which assure the truth of (5) will be introduced successively.

5. **The finite stage.** Suppose first that $\sum f(p^k)^2/p^k$ converges for every fixed prime p, where k $(= 1, 2, \cdots)$ is the summation index. Then two differentiations of (4) at $u = 0$ show that, if μ_p' and μ_p'' are abbreviations for $(1 - p^{-1}) \sum f(p^k)/p^k$ and $(1 - p^{-1}) \sum f(p^k)^2/p^k$, the asymptotic distribution function, $\phi^p = \phi^p(\alpha)$, of the additive function $f^p(n)$ has μ_p' and μ_p'' as momenta of first and second order respectively. Hence, its standard deviation is $\lambda_p^{\frac{1}{2}}$, where $\lambda_p = \mu_p'' - \mu_p'^2$. Thus, if $p, r, \cdots$ is any finite set of primes, the standard deviation of the convolution $\phi^p * \phi^r * \cdots$ follows by substituting $\lambda_p = \mu_p'' - \mu_p'^2, \lambda_r = \mu_r'' - \mu_r'^2, \cdots$ into $(\lambda_p + \lambda_r + \cdots)^{\frac{1}{2}}$. But, if $p, r, \cdots$ are distinct, then, as mentioned before (4), the additive functions $f^p(n), f^r(n), \cdots$ of n are statistically independent (no matter how the additive function $f(n)$ be chosen). Consequently,[4] the function $f^p(n) + f^r(n) + \cdots$ of n has an asymptotic distribution function which is precisely the convolution $\phi^p * \phi^r * \cdots$ of the asymptotic distribution functions of the functions $f^p(n), f^r(n), \cdots$ of n. Accordingly (see footnote 4), if $f^p(n) + f^r(n) + \cdots$ is chosen to be a partial sum of the series $f(n) = \sum f^p(n)$, say the partial sum

$$\sum_{p<x} f^p(n),$$

then

$$\sum_{p<x.} \lambda_p \equiv \sum_{p<x} \left\{(1 - p^{-1}) \sum_{k=1}^{\infty} f(p^k)^2/p^k - (1 - p^{-1})^2 \left(\sum_{k=1}^{\infty} f(p^k)/p^k\right)^2\right\}$$

is the square of the standard deviation of the asymptotic distribution function of the additive function represented by the partial sum. Let the square of the latter standard deviation, which is a function of x, be denoted by $\lambda(x)$.

Since $\lambda(x)$ is the multiple sum in the last formula line, it is clear that, if the (arbitrary) values $f(p^k)$ defining the additive function $f(n)$ do not increase too rapidly as $p \to \infty$, $k \to \infty$ (for instance, if

$$|f(p^k)| < Ck^c \tag{6}$$

holds for two sufficiently large constants C, c), then

$$\lambda(x) - \sum_{p<x} \left[\frac{(1 - p^{-1})f(p^2)}{p} - (1 - p^{-1})^2 \left(\frac{f(p)}{p}\right)^2\right]$$

tends to a limit as $x \to \infty$ (the last sum being the contribution of $k = 1$ alone). Finally, if the values $f(p^k)$ belonging to $k = 1$ are so chosen that $f(p) = 1$ for

[4] Cf. B. Jessen and A. Wintner, *Distribution functions and the Riemann zeta function*, Transactions of the American Mathematical Society, vol. 38(1935), pp. 48–88, more particularly, pp. 84–85 and 56–57.

every p, then the last sum, $\sum [\ ]$, reduces to $\log\log x + \text{const.} + o(1)$ as $x \to \infty$ (Mertens).[5]

Thus, if (6) is satisfied and $f(p) = 1$, then $\lambda(x) - \log\log x$ tends to a finite limit as $x \to \infty$. In particular, $\lambda(x) \sim \log\log x$. Consequently, the Poissonian formula (5) can now be written in the form

$$f_m(x) \sim L_m(x) \tag{7}$$

as $x \to \infty$, where

$$L_m(x) = \frac{x(\log x)^{-1}(\log\log x)^{m-1}}{(m-1)!}.$$

Accordingly, what remains to be established is a set of conditions which, in conjunction with (6) and $f(p) = 1$, assure the truth of (7) for the additive function $f(n)$ assigned by the values $f(p^k)$; it is understood that m is arbitrarily fixed.

6. **The Tauberian condition.** It will now be proved that such a set of conditions is represented by the positivity of all the values $f(p^k)$. What will actually be shown is that (7) holds for every m whenever the conditions

(8a) $f(p) = 1,$ (8b) $f(p^k) > 0$ $(k = 2, 3, 4, \cdots)$

are satisfied for every prime p so that no limitation of the type (6) will now be needed. It is clear that, while (8a) is the *homogeneity* condition mentioned after (4), condition (8b) represents the *Tauberian* restriction referred to after (5).

It is instructive that this form of a Tauberian restriction is so essential as to become inadequate even in the limiting case, where the inequalities (8b) become equalities. In fact, let $f(n)$ be that additive function for which $f(p^k)$ is 1 or 0 according as $k = 1$ or $k > 1$ (cf. (8a) and (8b), where p is arbitrary). Clearly, $f(n)$ then is the number of the *simple* prime factors of n so that $f(n)$ is the function $\theta(n)$ defined before (2). Hence, (2) shows that the Poissonian law (7) is false in this case (although even (6) is satisfied). The simplest instances of additive functions $f(n)$ satisfying all the conditions (8a), (8b), (6) are two classical functions,[6] the first of which represents the number of the *distinct* prime divisors of n and the second the number of *all* prime divisors of n. In fact, it is clear that these two additive functions $f(n)$ are respectively defined by $f(p^k) = 1$ and $f(p^k) = k$, where p and k are arbitrary. In view of these examples, it is worth emphasizing that the values $f(p^k)$ will not be restricted to integers.

7. **A lemma.** The proof of the fact that (8a) and (8b) imply (7) can be based on a remark which has nothing to do with additive functions, since it can be formulated as follows.

Let $\theta(n)$ denote the function defined in (2), *and let S be any set of positive integers which has the property that there exists a fixed positive integer $m = m(S)$ satisfying*

[5] Cf. C. F. Gauss, loc. cit. (see footnote 1), p. 12 and p. 17.

[6] Cf. E. Landau, loc. cit. (see footnote 1).

the following conditions: No positive integer n satisfying $\theta(n) > m$ is in S and a positive integer n satisfying $\theta(n) = m$ is or is not in S according as n is or is not square-free. Then

$$S(x) \sim L_m(x) \tag{9}$$

as $x \to \infty$, where $L_m(x)$ is the same function as in (1) *and $S(x)$ denotes the number of those elements of S which are less than x.*

The integers n about which it is not assumed whether or not they are in S are those integers n for which neither $\theta(n) > m$ nor $\theta(n) = m$; hence $\theta(n) < m$, and so *these* integers n are in one of the $m - 1$ classes defined by $\theta(n) = 1, \theta(n) = 2, \cdots, \theta(n) = m - 1$. But $\theta_1(x), \theta_2(x), \cdots, \theta_{m-1}(x)$ denote the number of *all* integers n less than x satisfying $\theta(n) = 1, \theta(n) = 2, \cdots, \theta(n) = m - 1$ respectively. Since (2) and (1) imply that $\theta_1(x) + \cdots + \theta_{m-1}(x) = o(L_m(x))$, it follows that the number of those elements of S which are less than x and satisfy neither $\theta(n) > m$ nor $\theta(n) = m$ is $o(L_m(x))$. On the other hand, an integer n satisfying either $\theta(n) > m$ or $\theta(n) = m$ is supposed to be in S if and only if n satisfies $\theta(n) = m$ and is square-free. Since $\theta(n)$ denotes the number of simple prime factors of n, it follows that the number of those elements of S less than x which have not been enumerated by the preceding $o(L_m(x))$ is identical with the number of those positive integers less than x which are composed of m simple prime factors. But the latter number is $\pi_m(x)$, by the definition of $\pi_m(x)$ in (1). Accordingly, $S(x)$ is the sum of $o(L_m(x))$ and $\pi_m(x)$. Hence, (9) follows from (1).

8. **The proof.** In order to deduce (7) from (9), let S be the set of those positive integers for which

$$S: \qquad f(n) = m, \tag{10}$$

where $f(n)$ is a fixed additive function and the integer m has a given value. Then the definition of $f_m(x)$ at the beginning of this paper shows that (7) is identical with (9). Hence, all that remains to be shown is that, if (8a) and (8b) are satisfied, the n-set S defined by (10) fulfills the conditions under which (9) has been established.

To this end, let $p_1 = p_1(n), p_2 = p_2(n), \cdots$ denote the *simple* prime factors of an arbitrary positive integer n (> 1). Since $f(n)$ is additive, it is clear from (8b) that $f(n) \geqq f(p_1) + f(p_2) + \cdots$ according as not all or all prime factors of n are simple. Since (8a) and the definition of $\theta(n)$ imply that $f(p_1) + f(p_2) + \cdots = \theta(n)$, it follows that $f(n) > \theta(n)$ or $f(n) = \theta(n)$ according as n is not or is square-free. It follows that no n satisfying $\theta(n) > m$ is in the set (10) and that an n satisfying $\theta(n) = m$ is in the set (10) if and only if n is square-free. Since these properties are exactly the properties required of S in (9), the proof is complete.

THE JOHNS HOPKINS UNIVERSITY.

PARAMETRIC SOLUTIONS OF CERTAIN DIOPHANTINE EQUATIONS

By E. T. Bell

1. **Introduction.** The complete integer solution of

$$(1.1)\qquad x_1y_1 + \cdots + x_ny_n = 0$$

is given by the formulas[1]

$$(1.2)\qquad x_i = \alpha\alpha_i, \qquad y_i = -\sum_{j=1}^{i-1}\alpha_j\beta_{j,i} + \sum_{j=1}^{n-i}\alpha_{i+j}\beta_{i,i+j},$$

where the Greek letters denote integer parameters, with the convention (as always) that a summation (or range of values) in which the lower limit exceeds the upper is vacuous. Let $f_i(x_1, \cdots, x_n)$ $(i = 1, \cdots, n)$ be any functions which for integer values of $x_1, \cdots, x_n$ take integer values. Then the transformation

$$(1.3)\qquad y_i \to y_i - f_i(x_1, \cdots, x_n) \qquad (i = 1, \cdots, n)$$

takes (1.1) into

$$(1.4)\qquad \sum_{i=1}^{n} x_if_i(x_1, \cdots, x_n) = \sum_{i=1}^{n} x_iy_i,$$

and the complete integer solution of (1.4) is

$$(1.5)\qquad \begin{aligned} &x_i = \alpha\alpha_i, \\ &y_i = -\sum_{j=1}^{i-1}\alpha_i\beta_{j,i} + \sum_{j=1}^{n-i}\alpha_{i+j}\beta_{i,i+j} + f_i(\alpha\alpha_1, \cdots, \alpha\alpha_n). \end{aligned}$$

These solutions are valid in any Euclidean ring, as may be seen from the proof (see footnote 1) of (1.2). The like, therefore, holds for equations and their solutions obtained from (1.1), (1.4) by operating within any given Euclidean ring.

Equations of the types (1.6)–(1.11) are to be considered.

$$(1.6)\qquad a_1x_1\cdots x_{i_1} + a_2y_1\cdots y_{i_2} + \cdots + a_nz_1\cdots z_{i_n} = 0,$$

in which $a_1, \cdots, a_n$ are constant integers $\neq 0$ and $i_1 > 1, i_2 > 1, \cdots, i_n > 1$. This is one possible generalization of (1.1); its complete integer solution proceeds from (1.2).

$$(1.7)\qquad \sum_{j=1}^{n} a_jx_{j1}\cdots x_{ji_j}f_j(x_{j1}, \cdots, x_{ji_j}) = \sum_{j=1}^{n} a_jx_{j1}\cdots x_{ji_j}y_j,$$

Received March 5, 1942.

[1] Th. Skolem, *Diophantische Gleichungen*, 1938, p. 20. The form of this solution is considerably simpler than that given by the method of L. Aubry, *Réponse á la solution générale par identités de l'équation par V. G. Tariste*, L'Intermédiare des Mathématiciens, vol. 23 (1916), pp. 133–134, reproduced in Dickson (see footnote 2), p. 194.

in which the a_j are constant integers $\neq 0$ and the f_j are as in (1.4). The complete integer solution of (1.7) follows from that of (1.6) in the same way that (1.5) follows from (1.2), so that (1.7) need not be further discussed.

$$Q(x_1, \cdots, x_n) = x_1y_1 + \cdots + x_ny_n, \tag{1.8}$$

where Q is the general homogeneous quadratic form in $x_1, \cdots, x_n$ with integer coefficients. The complete integer solution of (1.8) is obtained.

$$Q(x_1, \cdots, x_{n-1}) = uv, \tag{1.9}$$

where Q is the general homogeneous quadratic form in $x_1, \cdots, x_{n-1}$ with integer coefficients.

The complete integer solution of (1.9) with $n - 1 = 2$ was found by Dickson,[2] using the classical theory of binary quadratic forms. Using his generalized quaternions, Dickson (see footnote 2, p. 193) found the complete integer solution of

$$x_1^2 - ax_2^2 - bx_3^2 + abx_4^2 = uv$$

for certain special values of a, b, and a parametric solution for arbitrary integers a, b. Latimer[3] obtained more general results for the same equation. In the lack of complete solutions for (1.9) with $n > 3$, it may be of interest to record a parametric solution, although there is no reason to suppose that this is the complete integer solution for any particular value of n, even if for some n the parametric solution of special cases of (1.9) can be identified with complete solutions found otherwise.[4]

$$Q(x_1, \cdots, x_n) = a_1x_1y_1 + \cdots + a_nx_ny_n, \tag{1.10}$$

where Q is as in (1.8) and $a_1, \cdots, a_n$ are constant integers different from zero.

$$Q(x_1, \cdots, x_{n-1}) = auv, \tag{1.11}$$

which differs from (1.9) only by the constant coefficient $a \neq 0$.

2. **Equation** (1.6). The solution of an equation of this type is reduced to that of an equation of type (1.1) and an associated multiplicative system. The first

[2] L. E. Dickson, *Modern Elementary Theory of Numbers*, University of Chicago, 1939, p. 190.

[3] C. G. Latimer's extensions are summarized in Dickson (see footnote 2, p. 193).

[4] The parametric solution of (1.9) is obtained by determining certain of the parameters in the complete integer solution of another equation, say E, so that some of the indeterminates vanish. By suitable choice of the coefficients, E then degenerates to (1.9). Simultaneously the complete integer solution of E degenerates to a parametric solution of (1.9). There are an infinity of equations which degenerate in this way to (1.9), and each furnishes a parametric solution of (1.9). If the entire class of equations which degenerate to a given equation could be defined, something might be inferred about the completeness of the degenerated solution. But the class appears to be undefinable constructively.

term in (1.6) is written $a_1x_1 \cdots x_{i_1-1} \times x_{i_1}$, and similarly for all. Then, by (1.1), (1.2),

$$a_1x_1 \cdots x_{i_1-1} = \alpha\alpha_1, \quad a_2y_1 \cdots y_{i_2-1} = \alpha\alpha_2, \quad \cdots, \quad a_nz_1 \cdots z_{i_n-1} = \alpha\alpha_n,$$

the associated multiplicative system, to be solved for the $x, y, \cdots, z$ and the α. The solution is given non-tentatively by either of two methods (see footnote 1, Skolem, Kap. 4) and the values thus found for $\alpha_1, \cdots, \alpha_n$ are then substituted in the expressions for $x_{i_1}, y_{i_2}, \cdots, z_{i_n}$ obtained from the second set of equations in (1.2) by writing $x_{i_1}, y_{i_2}, \cdots, z_{i_n}$ for $y_1, y_2, \cdots, y_n$ respectively.

The complete solution falls into classes of solutions, each class containing a single representative, according to the divisors of the constant coefficients $a_1, \cdots, a_n$ in (1.6). We discuss only the case in which the Euclidean ring concerned is that of the rational integers.[5]

Let $p_1, p_2, p_3, \cdots$ be the primes $2, 3, 5, \cdots$ in ascending order. The positive integer m may be written $m = p_\xi^{m_\xi}$, where the repeated Greek suffix indicates a product convention (analagous to the sum convention in tensors) over $\xi = 1, 2, 3, \cdots$; thus $p_\xi^{m_\xi} \equiv \prod_{s=1}^{\infty} p_s^{m_s}$. If p_t is the greatest prime dividing m, $m_s = 0$ for $s > t$.

In solving (1.6), the constants a_i are temporarily replaced by distinct indeterminates, $a_i \to u_i$, and the resulting equation, with all coefficients 1, is solved. Let $\theta_1, \cdots, \theta_r$ be all those parameters in the solution that appear in the parametric expressions of the u_i, so that $u_i = \theta_1^{c_{i1}} \cdots \theta_r^{c_{ir}}$, where $c_{i1}, \cdots, c_{ir}$ are integers $\geqq 0$. Hence

$$a_i = \theta_1^{c_{i1}} \cdots \theta_r^{c_{ir}} \qquad (i = 1, \cdots, n), \tag{2.1}$$

and at least one of $\theta_1, \cdots, \theta_r$ is a divisor of some a_i. All sets $(\theta_1, \cdots, \theta_r)$ satisfying (2.1) may be found (when $a_1, \cdots, a_n$ are any given integers) by solving a set (2.2) of linear Diophantine equations in non-negative integers. From (2.1),

$$\begin{aligned} p_\xi^{a_{i\xi}} &= (p_\xi^{\theta_{1\xi}})^{c_{i1}} \cdots (p_\xi^{\theta_{r\xi}})^{c_{ir}} \\ &= p_\xi^{c_{i1}\theta_{1\xi}+\cdots+c_{ir}\theta_{r\xi}} \qquad (i = 1, \cdots, n), \end{aligned}$$

and, therefore,

$$c_{i1}\theta_{1\xi} + \cdots + c_{ir}\theta_{r\xi} = a_{i\xi} \qquad (i = 1, \cdots, n; \xi = 1, \cdots, t), \tag{2.2}$$

in which $a_{i\xi}$ and the $c_{ij} \geqq 0$ are given constant integers and p_t is the greatest prime dividing $a_1 \cdots a_n$. A solution $(\theta_{1\xi}, \cdots, \theta_{r\xi}) = (\theta'_{1\xi}, \cdots, \theta'_{r\xi})$ of (2.2) determines a set of the constants appearing as coefficients in the complete solution of (1.6), which is separated into classes according to the solutions of (2.2). The determination of the number of classes is a solvable problem in compound partitions.

[5] A slight modification of the device employed for this case takes care of the general Euclidean ring.

3. **Equation** (1.8). The solution of this equation is a preliminary to that of (1.9). The general n-ary homogeneous quadratic form with integer coefficients is

$$Q(x_1, \cdots, x_n) \equiv \sum_{i=1}^{n} c_{i,i} x_i^2 + \sum_{i=1}^{n-1} \sum_{j=1}^{n-i} c_{i,i+j} x_i x_{i+j},$$

and the general linear transformation with integer coefficients which takes $x_1 y_1 + \cdots + x_n y_n$ into this is

$$y_i \to \sum_{j=1}^{i-1} (c_{j,i} - \gamma_{j,i})\, x_j + c_{i,i} x_i + \sum_{j=1}^{n-i} \gamma_{i,i+j} x_{i+j},$$

where the γ's are $\frac{1}{2}n(n-1)$ independent integer parameters. From (1.3) and (1.4), it follows that the complete integer solution of (1.8) is

$$y_i = \sum_{j=1}^{i-1} \alpha_j [\alpha(c_{j,i} - \gamma_{j,i}) - \beta_{j,i}] + c_{i,i}\alpha\alpha_i + \sum_{j=1}^{n-i} \alpha_{i+j}(\alpha\gamma_{i,i+j} + \beta_{i,i+j}),$$

$$x_i = \alpha\alpha_i \qquad (i = 1, \cdots, n).$$

There are $\frac{1}{2}n(n-1)$ parameters $\beta_{i,j}$ and $n+1$ parameters α, α_i. Hence the total number of parameters in the solution is $n^2 + 1$.

4. **Equation** (1.9). With the notation as in §3, and $i = 1, \cdots, n$, write

$$\delta_{i,j} \equiv \alpha(c_{j,i} - \gamma_{j,i}) - \beta_{j,i} \qquad (j = 1, \cdots, i-1),$$

$$\delta_{i,i} \equiv \alpha c_{i,i},$$

$$\delta_{i,i+j} \equiv \alpha\gamma_{i,i+j} + \beta_{i,i+j} \qquad (j = 1, \cdots, n-i).$$

Then

$$y_i = \sum_{j=1}^{n} \delta_{i,j}\alpha_j,$$

and $y_s = 0$ $(s = 1, \cdots, n-1)$ is the system

$$\sum_{j=1}^{n} \delta_{s,j}\alpha_j = 0 \qquad (s = 1, \cdots, n-1) \tag{4.1}$$

of $n-1$ homogeneous linear equations in $\alpha_1, \cdots, \alpha_n$ with the $(n-1) \times n$ matrix $(\delta_{s,j})$. Denote by $(\delta_{s,j})_i$ the determinant of order $n-1$ obtained from $(\delta_{s,j})$ by deleting the i-th column of $(\delta_{s,j})$, and let δ be the G.C.D. of $(\delta_{s,j})_1, \cdots, (\delta_{s,j})_n$. Let λ denote an integer parameter, and write $\mu \equiv \lambda/\delta$. Then the complete integer solution of (4.1) is

$$\alpha_i = (-1)^{i-1}\mu(\delta_{s,j})_i \qquad (i = 1, \cdots, n). \tag{4.2}$$

The values (4.2) of α_i substituted into the solution in §3, with Q as there, give a parametric solution of

$$Q(x_1, \cdots, x_n) = x_n y_n$$

as polynomials with integer coefficients in $n^2 - n + 2$ integer parameters. If in this equation and its solution all $c_{r,s}$ in which at least one of r, s is n are set equal to zero, and the notation x_n, y_n is changed to u, v, we get a parametric solution of (1.9). If D_n is the determinant of $y_1, \cdots, y_n$ considered as linear forms in $\alpha_1, \cdots, \alpha_n$, and D'_n is the determinant obtained from D_n by setting $c_{i,n} = c_{n,i} = 0$ $(i = 1, \cdots, n)$, the value of v is $(-1)^{n-1}\mu D'_n$.

5. **Equations** (1.10), (1.11). The form Q is as in §3; the first equation to be considered is

$$Q(x_1, \cdots, x_n) = a_1x_1y_1 + \cdots + a_nx_ny_n, \tag{1.10}$$

with $a_1 \cdots a_n \neq 0$. The method for (1.10) is an immediate extension of that for (1.9) in §3.

Necessary and sufficient conditions that $y_i \to y'_i$, where

$$y'_i \equiv \sum_{j=1}^{n} \gamma_{i,j}x_j \qquad (i = 1, \cdots, n) \tag{5.1}$$

with integer coefficients $\gamma_{i,j}$, take $\sum a_ix_iy_i$ into $Q(x_1, \cdots, x_n)$ are

$$a_i\gamma_{i,i} = c_{i,i}, \qquad a_i\gamma_{i,j} + a_j\gamma_{j,i} = c_{i,j} \qquad (i < j). \tag{5.2}$$

It will be assumed that these conditions are satisfied. Hence, if $a_{i,j}$ is the G.C.D. of a_i, a_j, it is necessary that $a_i \mid c_{i,i}$ and $a_{i,j} \mid c_{i,j}$, so that

$$c_{i,i} = a_ic'_{i,i}, \qquad a_i = a_{i,j}a'_i, \qquad a_j = a_{i,j}a'_j, \qquad c_{i,j} = a_{i,j}c'_{i,j},$$

where all the letters denote integers and a'_i, a'_j, $c'_{i,j}$ are coprime in pairs. Hence $\gamma_{i,i} = c'_{i,i}$. The general solution of the second equation (5.2) is

$$\gamma_{i,j} = c'_{i,j}\gamma'_{i,j} + a'_j\lambda_{i,j}, \qquad \gamma_{j,i} = c'_{i,j}\gamma'_{j,i} - a'_i\lambda_{i,j} \qquad (i < j),$$

where $\gamma'_{i,j}$, $\gamma'_{j,i}$ is any solution of

$$a'_i\gamma'_{i,j} + a'_j\gamma'_{j,i} = 1,$$

and $\lambda_{i,j}$ is an integer parameter. From (5.1) we have

$$y'_i = \sum_{j=1}^{i-1} (c'_{j,i}\gamma'_{i,j} - a'_j\lambda_{j,i})\, x_i + c'_{i,i}\, x_i + \sum_{j=1}^{n-i} (c'_{i,i+j}\gamma'_{i,i+j} + a'_{i+j}\lambda_{i,i+j})x_{i+j}.$$

Hence (1.10) may be written

$$\sum_{i=1}^{n} a_i\, x_i y''_i = 0, \qquad y''_i \equiv y_i - y'_i, \tag{5.3}$$

which is of type (1.6), the associated multiplicative system being

$$a_ix_i = \alpha\alpha_i \qquad (i = 1, \cdots, n). \tag{5.4}$$

From the solution x_i, y''_i of (5.3), the solution of (1.10) is written down in an obvious way. From this a parametric solution of (1.11) is obtained as in §4 for (1.9) from (1.8).

California Institute of Technology.

THE DOUBLE-N_n CONFIGURATION

By Arthur B. Coble

1. **Introduction.** We are concerned initially only with the double-N_n configuration defined by a White [**7**] surface, a configuration in the linear space $[n+1]$ which consists of N_n lines, l_i, and of N_n spaces $[n-1]$, λ_j, such that l_i and λ_j are incident if and only if $i \neq j$ $(i, j = 1, \cdots, N_n)$, where N_n is the binomial coefficient $(n+2; 2)$. In conclusion, however, we raise the question as to whether there may not well be configurations of this sort more general than those which define, and are defined by, a White surface.

The White surface, W_n, in $[n+1]$ is the map of the plane by the linear system of curves of order $n+1$ on a generic set, $P^2_{N_n}$, of points $p_1, \cdots, p_{N_n}$ of the plane. It has the order $(n+1)^2 - N_n = N_{n-1}$. The directions about a point p_i, say p_i^*, map into points on a line l_i of the configuration CW_n under discussion. Let C_j be the curve of order n on all of the points of $P^2_{N_n}$ except p_j. Since $P^2_{N_n}$ is generic, the curves C_j are all distinct and generic. A particular curve C_j maps into a curve k_j in an $[n-1]$ of order N_{n-2} which crosses each of the lines l_i $(i \neq j)$, since C_j goes through p_i with some definite direction. Let λ_j be the $[n-1]$ in which k_j lies. Then λ_j also cuts l_i if $i \neq j$. We thus obtain from $P^2_{N_n}$ the double-N_n configuration of the type we will call CW_n. The configuration itself, apart from the W_n which defines it, is formally self-dual. The lines and $[n-1]$'s are dual in $[n+1]$. The non-incidence of l_i and λ_i imply that they have neither a $[0]$ nor a $[n]$ in common. The incidence of l_i and λ_j $(j \neq i)$ imply that they have a point m_{ij} in common and a prime μ_{ij} in common.

The first instance for $n = 1$, $N_n = 3$ is the figure of three lines l_1, l_2, l_3 in the plane W_1 and three points λ_1, λ_2, λ_3, each point on two of the lines, i.e., a plane triangle. The mapping mentioned above by conics on P^2_3 is a quadratic transformation from the plane of P^2_3 to W_1. In this case the CW_1 has no geometric interest.

The second instance for $n = 2$, $N_n = 6$ is the figure of six skew lines $l_1, \cdots, l_6$ on a cubic surface W_2 in [3] and the six lines $\lambda_1, \cdots, \lambda_6$, which with $l_1, \cdots, l_6$ form a "double-six" on the surface. This figure has two significant properties for which we use the terms *descriptively self-dual* and *intrinsically self-dual* with the following meanings. A formally self-dual configuration is descriptively self-dual if for every figure constructed from its parts there exists a dual figure dually constructed from its dual parts. A descriptively self-dual configuration is intrinsically self-dual if there exists a correlation which transforms each part into its dual part. Naturally the formal, descriptive, and intrinsic

Received March 24, 1942. This and the following article present material which was reported in a retiring address delivered at the Dallas meeting of the American Association for the Advancement of Science.

self-duality are successive stages of restriction each implying the existence of the preceding stages. The CW_2 is an example of intrinsic self-duality. An instance of its descriptive self-duality is the following dual theorem: The points [primes] of the lines $l_1, \cdots, l_6$ [$\lambda_1, \cdots, \lambda_6$] are on a unique cubic surface W_2 [envelope $\overline{W}_2$] which also contains the points [primes] of the lines $\lambda_1, \cdots, \lambda_6$ [$l_1, \cdots, l_6$]. Furthermore the existence of the Schur [6; 5] quadric attached to the double-six, in which the lines l_i, λ_i are pole and polar, shows that CW_2 is also intrinsically self-dual. It is clear of course that the correlation which exists in the case of an intrinsically self-dual configuration C is unique and involutorial unless C itself is so special as to admit collineations.

It is the purpose of this article to show that, for $n > 2$, the CW_n is not descriptively self-dual for generic $P^2_{N_n}$; that when descriptive self-duality is restored, in some measure at least, by restricting $P^2_{N_n}$, then intrinsic self-duality does not yet exist; and that finally there exists a class of sets $P^2_{N_n}$ with $2n + 2$ absolute constants for which the CW_n is intrinsically self-dual. That CW_n for generic $P^2_{N_n}$ is not intrinsically self-dual is proved by Room [**5**;77].

One is always inclined to accept as inevitable that attenuation of geometric properties which accompanies the process of generalization to spaces of higher dimension. The basis of the attenuation is the greater variety of possibilities in the higher dimension. Thus a planar conic generalizes into either a cubic space curve or into a quadric surface and the properties of the conic are dispersed over these two diverse figures. It is interesting then to find instances of the fact that all the salient features of a figure may be preserved under generalization provided the generalization is followed by properly chosen specialization.

2. **Non-duality of the generic CW_n.** We wish to prove the theorem:

(1) *For $n \geqq 3$ the CW_n derived from a generic set $P^2_{N_n}$ is not a descriptively self-dual figure.*

For this purpose we have to show that CW_n defines projectively a figure such that the dual figure dually defined does not exist. Let $P^2_{N_n}$ be generic, and let W_n be the White surface of points obtained by mapping the plane upon W_n as described above. We consider first the prime μ_{ij} on the incident line, l_i, and $[n-1]$, λ_j. This prime corresponds to a curve H_{ij} of the mapping system which has a node at p_i and the factor C_j, since μ_{ij} cuts W_n in l_i and the curve k_j. Hence H_{ij} is the product of the line (p_ip_j) and C_j, and the prime μ_{ij} cuts W_n in l_i, k_j, and a further residual rational norm-curve N_{ij} of order $n - 1$, the map of the line (p_ip_j). In the pencil π_j of primes on λ_j, the $N_n - 1$ primes μ_{ij} $(i \neq j)$ are projective to the $N_n - 1$ lines (p_ip_j) of the pencil on p_j, since the curves H_{ij} have the fixed factor C_j and the variable factor (p_ip_j). Hence

(2) *The $N_n(N_n - 1)$ primes μ_{ij} are arranged in N_n pencils π_j which projectively are the pencils of a ternary N_n-point, $P^2_{N_n}$.*

The algebraic conditions that pencils π_1, π_2, π_3 be those of a ternary n-point are of the following simple form [**1**; 195, (14)]:

$$D(1;\ 23,\ 45)\cdot D(2;\ 31,\ 45)\cdot D(3;\ 12,\ 45) = 1, \tag{3}$$

where $D(1; 23, 45)$ is the double ratio of $\mu_{12}\mu_{13}$, $\mu_{14}\mu_{15}$.

The dual of the primes μ_{ij} are the points m_{ij}, where the line l_i is met by λ_j. These points m_{ij} are also distributed in N_n pencils ρ_i on the lines l_i. With respect to these we prove the theorem that

(4) *The N_n pencils ρ_i of points m_{ij} on lines l_i of CW_n are not projective to the pencils of a ternary N_n-point $P'^2_{N_n}$ when $n > 2$, and when the $P^2_{N_n}$ which defines CW_n is generic.*

The proof of this theorem carries with it the proof of (1) since (2) asserts the existence of a projective property and (4) asserts the non-existence of the dual property.

The point m_{ij} on W_n is on l_i and k_j. It is therefore the map of the direction at p_i on C_j. Thus the pencil ρ_i of points m_{ij} is projective to the pencil of tangents to curves C_j at p_i.

For $n = 2$ the opposite of (4) is correct, $P'^2_{N_n}$ being the same as $P^2_{N_n}$. Indeed, in this case the mapping is carried out by cubic curves on P^2_6. There is a pencil of such curves, nodal at p_i, and the pairs of nodal tangents are in an involution. Included among these nodal curves are the degenerate members $(p_ip_j)\cdot C_j$ whose tangents are (p_ip_j) and the tangent to C_j at p_j. Thus the pencil π_i on λ_i and the pencil ρ_i on l_i are projective in such wise that μ_{ji} corresponds to m_{ij}. This indeed is a consequence of the existence of the Schur quadric. This polarity interchanges l_i, λ_i and l_j, λ_j, and therefore also interchanges m_{ij} with μ_{ji}.

We next examine $n = 3$ to see that (4) is valid in that case. We remark provisionally that, if the pencils concerned belong to a P'^2_{10} for generic P^2_{10}, they must also do so for a particular P^2_{10}. For, the necessary double-ratio relations will continue to exist as the points of P^2_{10} change. Let, then, $p_1, \cdots, p_6$ be generically chosen, and let $p_7, \cdots, p_{10}$ be on a line ξ. The cubics $C_1, \cdots, C_6$ must contain the factor ξ with a residual factor which is a conic $c_1, \cdots, c_6$ on the five of these six points other than $p_1, \cdots, p_6$ respectively. Then, as for $n = 2$, the tangents to $C_2, \cdots, C_6$ at p_1 are those of $c_2, \cdots, c_6$ at p_1, and they are projective to the lines from p_1 to $p_2, \cdots, p_6$ under an involution I_1. Thus, however the remaining four points of the desired P'^2_{10} may lie, we may take the first six points of P'^2_{10} at $p_1, \cdots, p_6$. Let $\xi^{(1)}$, $\xi^{(2)}$, $\xi^{(3)}$ be lines on p_1, p_2, p_3 respectively which do *not* meet in a point; let $\xi^{(1)'}$, $\xi^{(2)'}$, $\xi^{(3)'}$ be their partners in the involutions I_1, I_2, I_3 respectively. Let C_7 be the cubic curve on $p_1, \cdots, p_6$ which has at p_1, p_2, p_3 the respective tangents $\xi^{(1)'}$, $\xi^{(2)'}$, $\xi^{(3)'}$; let ξ be an arbitrary line which meets C_7 in p_8, p_9, p_{10}; and let p_7 be a further arbitrary point on ξ. We seek to construct the point p'_7 of the desired P'^2_{10}. For this we add the tangents $\xi^{(1)'}$, $\xi^{(2)'}$, $\xi^{(3)'}$ to the pencils at p_1, p_2, p_3; and we

apply the involutions I_1, I_2, I_3 respectively to them so as to obtain the first six points of P'^2_{10} at $p_1, \cdots, p_6$, and thus get lines $\xi^{(1)}$, $\xi^{(2)}$, $\xi^{(3)}$ which do *not* meet in the required point p'_7 of the required P'^2_{10}. Since P'^2_{10} does not exist for this particular P^2_{10}, it does not exist for the generic P^2_{10}.

We complete the proof of (4) by showing that, if (4) is true for $n - 1$, it is also true for n. Let $P^2_{N_n}$ be chosen so that it is composed first of a generic set $P^2_{N_{n-1}}$ and a set of $n + 1$ remaining points on a line ξ. Then $C_1, \cdots, C_{N_{n-1}}$ have a common factor ξ and residual factors $c_1, \cdots, c_{N_{n-1}}$, which are the corresponding curves for the set $P^2_{N_{n-1}}$. If the tangents to the curves C of $P^2_{N_n}$ define a ternary set $P'^2_{N_n}$, it is necessary that the tangents to the curves c of $P^2_{N_{n-1}}$ define a set $P'^2_{N_{n-1}}$ which is a part of $P'^2_{N_n}$. Thus (4) can be true for n only if it is true for $n - 1$.

If we take account of the possibility of mapping the points of a plane upon either the points, or the primes, of $[n + 1]$, we may state the above results as follows.

(5) *For the generic CW_n with N_n lines l and N_n spaces λ, there must exist a White surface W_n which contains the points on the lines l, or else there must exist a White envelope $\overline{W}_n$ which contains the primes on the spaces λ, but W_n and $\overline{W}_n$ do not coexist.*

Because of this curious lack of symmetry in the White double-N configuration we raise the question in the final section as to whether configurations may exist which have neither a W_n nor a $\overline{W}_n$. In the next section we find a special class of cases for which both W_n and $\overline{W}_n$ exist.

3. **Configurations CW_n defined by a surface W_n and an envelope $\overline{W}_n$.** It is convenient to introduce the W_n analytically in terms of the trilinear form [**4**]

$$(1)\qquad T(2, n, n + 1) = (\alpha x)(\beta y)(\gamma z) = \sum\nolimits_{ijk} a_{ijk}x_iy_jz_k \qquad (i = 0, 1, 2; j = 0, \cdots, n; k = 0, \cdots, n + 1),$$

with ξ, η, ζ as contragredient coordinates in the spaces [2], $[n]$, $[n + 1]$ of x, y, z respectively. If $z^{(0)}$ on W_n is the map of $x^{(0)}$ on [2], the equation of $z^{(0)}$ in $[n + 1]$ is

$$(2)\qquad (z^{(0)}\zeta) = \begin{vmatrix} (\alpha x^{(0)})\,\beta_j\gamma_k \\ \zeta_k \end{vmatrix} = 0.$$

For each section, $Q^{N_n-1}(\zeta)$, of W_n by a given prime ζ, this equation (2) yields the corresponding curve, $Q^{n+1}(\zeta)$, of the mapping system on $P^2_{N_n}$. This set of points p_h is obtained from the pairs, $x, y = p_h, q_h$, which are neutral for z in $T = 0$, i.e.,

$$(3)\qquad (\alpha p_h)(\beta q_h)\gamma_k = 0.$$

These neutral pairs are defined independently of z by the double identity (cf. [4; §5, (4)]):

$$\sum_h (q_h\eta)\cdot(p_h\xi)^{n-1} \equiv 0 \qquad (h = 1, \cdots, N_n). \tag{4}$$

If in $T = 0$ we restrict z to the prime $(\zeta z) = 0$, then $T = 0$ becomes a $T'(2, n, n) = 0$ with spaces $[x]$, $[y]$, $[z'] = \zeta$. Then the curve $Q^{n+1}(\zeta)$ in (2) is the locus of points $x^{(0)}$ for which the bilinear form in y, z' is singular. For given $x^{(0)}$ on this curve, the two singular points $y^{(0)}$, $z^{(0)}$ are furnished by the equation

$$\begin{vmatrix} (\alpha x^{(0)})\beta_j\gamma_k & \eta_j \\ \zeta_k & 0 \\ \zeta'_k & 0 \end{vmatrix} = (z^{(0)}\zeta')\cdot(y^{(0)}\eta). \tag{5}$$

Thus $x^{(0)}$ on $Q^{n+1}(\zeta)$ is mapped upon $z^{(0)}$ on $Q^{N_n-1}(\zeta)$, and upon $y^{(0)}$ on $S^{N_n-1}(\zeta)$ by means of two complete linear series $g_n^{N_n-1}$, $g_n'^{N_n-1}$ which are residual to each other in the linear series cut out on $Q^{n+1}(\zeta)$ by all curves of order n. We call these two paired curves, $Q^{N_n-1}(\zeta)$, $S^{N_n-1}(\zeta)$, "Reye curves". The equivalences which define them projectively for given $Q^{n+1}(\zeta)$ on $P^2_{N_n}$ are as follows:

$$P^2_{N_n} + g_n^{N_n-1} \equiv (n+1)L, \qquad g_n^{N_n-1} + g_n'^{N_n-1} \equiv nL, \tag{6}$$

where L is a line section of $Q^{n+1}(\zeta)$. In the mapping (5), the points x_h of $P^2_{N_n}$ pass into the points z_h on W_n, where ζ cuts the lines l_h of CW_n, and, in the space $[y]$, into the points q_h of (3) and (4), a set which we call $Q^n_{N_n}$.

A pair of Reye curves are not in general projective to each other in such wise that $y^{(0)}$ corresponds to $z^{(0)}$. When this happens, and when the space $[n]$ of ζ in $T'(2, n, n)$ is projected upon the space $[y]$ so that $z^{(0)}$ falls on $y^{(0)}$, the form $T'(2, n, n)$ reduces to the polarized form $(\alpha'x)(\beta'y)^2_{z'}$. Then either Reye curve is a "Jacobian curve", the locus of nodes of quadrics of a net. The generic $T'(2, n, n)$ has $3(n+1)^2 - 9 - 2n(n+2) \doteq (n+1)^2 - 7$ absolute constants; the net of quadrics has $3N_n - 9 - n(n+2)$ absolute constants. Thus it is N_{n-2} conditions that a Reye curve becomes a Jacobian curve.

For the Jacobian curve, the linear series $g_n^{N_n-1}$, $g_n'^{N_n-1}$ on $Q^{n+1}(\zeta)$ coincide into a contact linear series. Since the generic $Q^{n+1}(\zeta)$ has the genus N_{n-2}, there are $\infty^{N_{n-2}}$ distinct series $g_n^{N_n-1}$, whereas there are only a finite number of contact $g_n^{N_n-1}$'s. This confirms the above number, N_{n-2}, of conditions for a Jacobian curve. In the contact case the equivalences (6) yield $2g_n^{N_n-1} \equiv nL$, $2P^2_{N_n} \equiv (n+2)L$. Moreover, this last equivalence, in combination with (6), which define $g_n^{N_n-1}$ and $g_n'^{N_n-1}$, respectively, yields again $g_n^{N_n-1} \equiv g_n'^{N_n-1}$. Hence, we have the following.

(7) *If $P^2_{N_n}$ on $Q^{n+1}(\zeta)$ satisfies the equivalence $2P^2_{N_n} \equiv (n+2)L$, then the Reye curves, $Q^{N_n-1}(\zeta)$, $S^{N_n-1}(\zeta)$, are projectively equivalent Jacobian curves.*

For given $P^2_{N_n}$, there are only ∞^{n+1} incident curves $Q^{n+1}(\zeta)$. Thus the N_{n-2} conditions that $2P^2_{N_n} \equiv (n+2)L$ cannot be satisfied if $n > 3$; they can be satisfied in ∞^1 ways if $n = 3$; and in ∞^2 ways if $n = 2$. We have, therefore, the theorem:

(8) *The generic White surface W_n has no prime sections which are Jacobian curves if $n \geqq 4$; and it has only ∞^1 such sections if $n = 3$. However, when $n = 2$ and W_2 is a cubic surface with isolated double-six, there will be ∞^2 sections ζ of W_2 for which the cubic curves $Q^3(\zeta)$ on P^2_6 are such that $2P^2_6 \equiv 4L$, and these sections ζ are the cubic envelope $\overline{W}_2$ which has the same double-six, CW_2.*

This special situation in connection with W_2 occurs also in connection with W_n if the set $P^2_{N_n}$ is the set of nodes of a rational plane curve, $\rho_2^{n+3}(t)$. We recall [**2**; §6] that if, on a plane $[x']$, the point x' is determined as the intersection of tangents t_1, t_2 of a norm-conic, $K(t)$, then the pairs of nodal parameters t_{1h}, t_{2h} at the nodes p_h of $\rho_2^{n+3}(t)$ determine in $[x']$ the "nodular" points p'_h of a nodular set $P'^2_{N_n}$. We recall also that in addition to the trilinear form $T(2, n, n+1)$ used above with pairs $x, y = p_h$, q_h neutral for z, there now exists also a second trilinear form $T'(2, n, n+1)$ in variables x', y, z' with pairs p'_h, q_h neutral for z'. For generic rational curve the nodal set $P^2_{N_n}$ and the nodular set $P'^2_{N_n}$ are not projectively equivalent. We prove the following theorem which includes the case W_2 above.

(9) *If $P^2_{N_n}$ is the set of nodes of a generic rational planar curve $\rho_2^{n+3}(t)$ [$3n$ absolute constants], and if the plane is mapped on the points of a White surface W_n in $[n+1]$, then $\rho_2^{n+3}(t)$ is mapped on a rational norm-curve $N^{n+1}(t)$ bisecant to the lines l_i of the CW_n. The ∞^2 collinear $(n+1)$-points of $\rho_2^{n+3}(t)$ are mapped on $(n+1)$-points of N^{n+1} on primes ζ which cut W_n in Jacobian curves. The ∞^2 primes thus obtained lie on a White envelope $\overline{W}_n$ which has the same CW_n as W_n. The primes of $\overline{W}_n$ are the map of points x' by curves of order $n+1$ on the nodular set $P'^2_{N_n}$.*

We observe first that the existence of $N^{n+1}(t)$ on W_n is, because of the mapping, an obvious consequence of the existence of the irreducible rational curve $\rho_2^{n+3}(t)$ with nodes at $P^2_{N_n}$. Let the line ξ cut $\rho_2^{n+3}(t)$ in the points $t_1, t_2, t_3, \cdots, t_{n+3}$. On the $n+1$ points $t_3, \cdots, t_{n+3}$ of $\rho_2^{n+3}(t)$ there is one, and only one, curve $Q^{n+1}(\zeta)$, since $\rho_2^{n+3}(t)$ is irreducible. Let L be the set of these $n+1$ points on $Q^{n+1}(\zeta)$. Using on $Q^{n+1}(\zeta)$ the equivalence $2P^2_{N_n} + L \equiv (n+3)L$, obtained from $\rho_2^{n+3}(t)$, we obtain the equivalence $2P^2_{N_n} \equiv (n+2)L$, which, according to (7), implies that the section of W_n by ζ, $Q^{N_n-1}(\zeta)$, as well as $S^{N_n-1}(\zeta)$, is a Jacobian curve. We have then only to show that the primes ζ so obtained are those of a White envelope $\overline{W}_n$ obtained by mapping from the plane $[x']$ by using $P'^2_{N_n}$.

We now introduce specific mappings to prove the remaining statements. We had obtained in [**2**; §9, (8)] from $T(2, n, n+1)$ a form,

$$\begin{aligned} &H(x^{n+1}, t^{n+1}) \\ (10)\qquad &= H_0(x)t^{n+1} + \binom{n+1}{1}H_1(x)t^n + \binom{n+2}{2}H_2(x)t^{n-1} + \cdots + H_{n+1}(x), \end{aligned}$$

whose polarized form $H(x^{n+1}, t_1, t_2, \cdots, t^{n+1}) = 0$ is the $(n+1)$-ic curve on the nodes $P^2_{N_n}$ of $\rho_2^{n+3}(t)$ and on the further points of $\rho_2^{n+3}(t)$ with parameters $t = t_1, t_2, \cdots, t_{n+1}$. Thus the mapping of $[x]$ on W_n is accomplished by setting

$$(11)\quad W_n: z_0 = H_0(x), \quad z_1 = H_1(x), \quad z_2 = H_2(x), \quad \cdots, \quad z_{n+1} = H_{n+1}(x).$$

Furthermore, when x is at the point $t = s$ of $\rho_2^{n+3}(t)$, then $H(x^{n+1}, t^{n+1})$ becomes the perfect power $(ts)^{n+1} = (t-s)^{n+1}$. Hence the curve $N^{n+1}(t)$ on W_n is given by

$$(12)\quad N^{n+1}(t): z_0 = 1, \quad z_1 = -t, \quad z_2 = t^2, \quad \cdots, \quad z_{n+1} = (-1)^{n+1}t^{n+1}.$$

There is also a well-known symmetric involutorial form,

$$(13)\qquad (\delta_1 t_1)^{n+1}(\delta_2 t_2)^{n+1}(\delta t)^{n+1} = 0,$$

which expresses that the points of $\rho_2^{n+3}(t)$ with parameters t_1, t_2, t are collinear. If in this we replace the symmetric combinations of t_1, t_2 by coordinates x' referred to $K(t)$, it becomes a form

$$(14)\quad D(x'^{n+1}, t^{n+1}) = D_0(x')t^{n+1} + D_1(x')t^n + D_2(x')t^{n-1} + \cdots + D_{n+1}(x') = 0,$$

which furnishes for given $x' = x'(t_1, t_2)$ the parameters of the $n+1$ further intersections with $\rho_2^{n+3}(t)$ of the line joining the points t_1, t_2 of $\rho_2^{n+3}(t)$. If t_1, t_2 is the pair t_{1h}, t_{2h} of nodal parameters at the node p_h, then the $(n+1)$-ic in t of (13), (14) vanishes identically, or the curves $D_0(x') = 0, \cdots, D_{n+1}(x') = 0$ are on the nodular set $P'^2_{N_n}$. Assuming for the moment that the curves $D_k(x') = 0$ are linearly independent, a fact which will appear presently, we can use them to map the points x' upon the primes of $\overline{W}_n$ by the relations:

$$(15)\quad \overline{W}_n: \zeta_0 = D_{n+1}(x'), \quad \zeta_1 = -D_n(x'), \quad \cdots, \quad \zeta_{n+1} = (-1)^{n+1}D_0(x').$$

If $D(x'^{n+1}, t^{n+1}) = 0$ has roots $t = t_3, \cdots, t_{n+3}$, the parameters of the $n+1$ further intersections of the line joining the points t_1, t_2 of $\rho_2^{n+3}(t)$, and if $H(x^{n+1}; t_3, t_4, \cdots, t_{n+3}) = 0$ is the $(n+1)$-ic curve on $P^2_{N_n}$ which cuts $\rho_2^{n+3}(t)$ in the same points, then the apolarity condition of $D(x'^{n+1}, t^{n+1})$ and $H(x^{n+1}, t^{n+1})$ is the incidence condition of $z(x)$ on W_n with $\zeta(x')$ on $\overline{W}_n$, namely (cf. (11), (14)),

$$(16)\quad z_0\zeta_0 + z_1\zeta_1 + \cdots + z_{n+1}\zeta_{n+1} = H_0(x)\cdot D_{n+1}(x') - H_1(x)\cdot D_n(x') + \cdots = 0.$$

Let t_1, t_2 approach t_{1h}, t_{2h} in such a manner that the ratio $(t_{1h} - t_1)/(t_{2h} - t_2)$ approaches the value δ. Then the point $x'(t_1, t_2)$ approaches the nodular point p'_h in a certain direction, and the line joining t_1, t_2 of $\rho_2^{n+3}(t)$ is approaching a line ξ on the node p_h. The limiting position of the curve (16) which always passes through p_h is $\xi \cdot C_h$. Thus the primes of $\overline{W}_n$ which correspond to directions δ about p'_h are the primes on the $[n-1]$, λ_h of the CW_n. Thus CW_n and $C\overline{W}_n$ are identical. Moreover, these primes are enough to exhaust the space $[n+1]$ of ζ in (15) so that the curves D in (15) are linearly independent.

The case $n = 2$ for which there are ∞^2 rational curves $\rho_2^5(t)$ with nodes at P_6^2 deserves some notice. Of the ∞^3 cubic curves $Q^3(\zeta)$ on P_6^2 there are ∞^2 curves

$Q^3(\zeta)$ on which the equivalence $2P_6^2 \equiv 4L$ exists. These contribute the planes ζ of the envelope $\overline{W}_2$ which has the same double-six as the cubic surface W_2. The peculiarity is that any one of the ∞^2 collinear triads of a cubic $Q^3(\zeta)$ is also a collinear triad of some one of the ∞^2 rational curves $\rho_2^5(t)$. These collinear triads of $Q^3(\zeta)$ map into triads on the section of W_2 by ζ which are the contacts of a system of contact conics. For, the curves $\rho_2^5(t)$ map into cubic curves bisecant to the lines l_i; the lines ξ map into cubic curves bisecant to the lines λ_i. A cubic curve of one system meets a cubic curve of the other in five points, $z^{(1)}, \cdots, z^{(5)}$, and the two curves are on a quadric q which touches W_2 at these five points. Then the plane $(z^{(3)}, z^{(4)}, z^{(5)})$ is a plane ζ of $\overline{W}_2$ which cuts W_2 in a cubic curve, and q in a conic, which touch at $z^{(3)}, z^{(4)}, z^{(5)}$. Since the rational curves $\rho_2^5(t)$ have perspective conics, the case $n = 2$ will also appear in the next section.

The case $n = 3$ with P_{10}^2 nodal for a single $\rho_2^6(t)$ is also somewhat peculiar. For, then (compare [**2**; §8]) there is also a rational curve $\rho_2^6(\tau)$ with nodes at the nodular set $P_{10}'^2$. Thus, in addition to the curve $N^4(t)$ on W_3 bisecant to the lines l_h of CW_3, there is a dual envelope $N^4(\tau)$ on $\overline{W}_3$ with the planes λ_h of CW_3 as axes. This appears to be one instance of descriptive self-duality without the accompaniment of intrinsic self-duality though it is not possible to be categorical with respect to descriptive self-duality without examining all the associated loci of the figure to make sure their duals exist.

For cases $n > 3$ it does not seem likely that the existence of $N^{n+1}(t)$ on W_n implies the existence of a corresponding $N^{n+1}(\tau)$ on $\overline{W}_n$ so that the descriptive self-duality, restored for the points m_{ij} of CW_n as contrasted with the primes μ_{ij} of CW_n by assuming that $P_{N_n}^2$ is nodal, would fail to persist under a more searching examination. In the next paragraph we find a $P_{N_n}^2$ with $2n + 2$ absolute constants for which CW_n is intrinsically self-dual, being unaltered by a polarity.

4. **Intrinsically self-dual configurations,** CW_n. Two rational planar envelopes $r_2^k(t)$, $r_2^{n+3-k}(t)$ of class k and $n + 3 - k$ respectively, with lines in (1, 1) correspondence through like-named parameters t, "generate" a rational planar point-locus $\rho_2^{n+3}(t)$, unless they are specially situated in that some corresponding lines coincide throughout. Conversely, a given $\rho_2^{n+3}(t)$ can be generated in this way by means of curves, $r_2^k(t)$ and $r_2^{n+3-k}(t)$, "perspective" to it. We exclude the case $k = 1$ which implies that all of the nodes of $\rho_2^{n+3}(t)$ coincide at an $(n + 2)$-fold point, this being a perspective point, $r_2^1(t)$.

We consider in this section the curve $\rho_2^{n+3}(t)$ with a perspective conic, $K(t)$. If $\xi(t^{(0)})$ is a tangent of $K(t)$, then ξ is a perspective line of $K(t)$, and the tangents of $K(t)$ set up a (1, 1) correspondence between the points of ξ and of $\rho_2^{n+3}(t)$. Since, conversely, the class of the locus of lines joining corresponding points is only two, $n + 2$ of the corresponding pairs must coincide. Hence

(1) *The $n - 1$ conditions that a line ξ meet $\rho_2^{n+3}(t)$ in $n + 2$ points whose parameters on the line and on $\rho_2^{n+3}(t)$ are projective are poristic for the line ξ. Indeed,*

$n - 2$ of these conditions fall on $\rho_2^{n+3}(t)$ and imply that $\rho_2^{n+3}(t)$ has a perspective conic $K(t)$, and the remaining condition requires that ξ be a tangent of $K(t)$.

Thus the rational curve $\rho_2^{n+3}(t)$ under consideration, and also its nodal set $P_{N_n}^2$, has $3n - (n - 2) = 2n + 2$ absolute constants.

With a_i and b_j linear forms in x, let $\rho_2^{n+3}(t)$ be generated by its perspective conic $K(t)$ and a perspective envelope of class $n + 1$, given by

$$\begin{aligned} &a_0t^2 + a_1t + a_2 = 0, \\ &b_0t^{n+1} + b_1t^n + b_2t^{n-1} + \cdots + b_{n+1} = 0. \end{aligned} \tag{2}$$

The equation in variables x of $\rho_2^{n+3}(t)$ is the Sylvester resultant,

$$R = R_{n+1,2} = \begin{vmatrix} a_0 & a_1 & a_2 & 0 & \cdots & 0 \\ 0 & a_0 & a_1 & a_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & a_2 \\ b_0 & b_1 & b_2 & b_3 & \cdots & 0 \\ 0 & b_0 & b_1 & b_2 & \cdots & b_{n+1} \end{vmatrix} = 0. \tag{3}$$

The subscripts in $R_{n+1,2}$ indicate the number of rows of coefficients a, and of coefficients b, in R. We denote also by $R_{n+1-j,2-j}$ $(j = 0, 1, 2)$ the matrix obtained from R by dropping the last j rows of coefficients a, the last j rows of coefficients b, and the last j columns.

Two binary forms of orders m, μ have a unique apolar $(m + \mu - 2)$-ic. This, formed for the binary forms in (2), is the $(n + 1)$-ic:

$$\begin{aligned} H(x^{n+1}, t^{n+1}) &= \begin{vmatrix} R_{n,1} \\ 1, -\binom{n+1}{1}t, \binom{n+1}{2}t^2, \cdots, (-1)^{n+1}t^{n+1} \end{vmatrix} \\ &= H_0(x)t^{n+1} + \binom{n+1}{1}H_1(x)t^n + \cdots + H_{n+1}(x). \end{aligned} \tag{4}$$

If x is at the point $x(t')$ on $\rho_2^{n+3}(t)$, the forms (2) have a common root t' and the apolar form is $(tt')^{n+1}$. If x is at a node p_h of $\rho_2^{n+3}(t)$, the forms (2) have two common roots and the apolar form vanishes identically. Thus H is the form of (10) of §3 and $H(x^{n+1}, t_1, \cdots, t_{n+1}) = 0$ is the adjoint of order $n + 1$ of $\rho_2^{n+3}(t)$ which cuts $\rho_2^{n+3}(t)$ in further points $t = t_1, \cdots, t_{n+1}$.

If all the adjoints on x and x' also go through x'', then all the binary $(n + 1)$-ics apolar to $H(x^{n+1}, t^{n+1})$ and $H(x'^{n+1}, t^{n+1})$ are also apolar to $H(x''^{n+1}, t^{n+1})$. Let x'' be the point $x''(t_1, t_2)$, the join of tangents t_1 and t_2 of $K(t)$. Then the $(n + 1)$-ics apolar to $H(x''^{n+1}, t^{n+1})$ include $(a_0t^2 + a_1t + a_2)_{x''} = (tt_1)(tt_2)$ taken with an arbitrary $(n - 1)$-ic, $(\alpha t)^{n-1}$. This same system of $(n + 1)$-ics is apolar

to $H(x^{n+1}, t^{n+1})$, $H(x'^{n+1}, t^{n+1}) = (tt_1)^{n+1}, (tt_2)^{n+1}$ when x, x' are at the points t_1, t_2 of $\rho_2^{n+3}(t)$. Hence

(5) *The linear system, $H(x^{n+1}, t_1t_2t^{n-1}) = 0$, of adjoints of $\rho_2^{n+3}(t)$ on the points t_1, t_2 of $\rho_2^{n+3}(t)$, a system (∞^{n-1}) for variable t, has a further base point at the point $x(t_1, t_2)$ of intersection of tangents t_1, t_2 of $K(t)$.*

An immediate corollary of this is:

(6) *The linear system, $H(x^{n+1}, t_1, t_2, \cdots, t_kt^{n-k+1}) = 0$ of adjoints of $\rho_2^{n+3}(t)$ on the points $t_1, \cdots, t_k$ of $\rho_2^{n+3}(t)$, a system (∞^{n-k+1}) for variable t, has $\binom{k}{2}$ further base points at the intersections of the k tangents $t_1, \cdots, t_k$ of $K(t)$ $(k = 2, \cdots, n+1)$. Thus to pass through the $\binom{k}{2}$ intersections of such a circumscribed k-line of $K(t)$ imposes only k linear conditions on the adjoints of $\rho_2^{n+3}(t)$ when $k > 2$, and the linear system so defined has k further base points on $\rho_2^{n+3}(t)$.*

We now prove a theorem particularly important for our purpose, namely:

(7) *In the linear system (∞^{n+1}) of adjoints of $\rho_2^{n+3}(t)$ of order $n+1$, there is a linear system (∞^{n-1}) of adjoints with a node at p_h, a node of $\rho_2^{n+3}(t)$. In this latter system there is a linear system (∞^{n-3}) of adjoints with a triple point at p_h.*

Let t_{1h}, t_{2h} be the parameters of the node p_h on $\rho_2^{n+3}(t)$. Consider the behavior of the linear system (∞^{n-1}) of adjoints $H(x^{n+1}, t_1, t_2, t^{n-1}) = 0$ as t_1, t_2 approach t_{1h}, t_{2h}. Since the adjoints are already on p_h and pass through t_1, t_2 on $\rho_2^{n+3}(t)$, then, as t_1 approaches t_{1h}, the adjoint touches the branch t_{1h} of $\rho_2^{n+3}(t)$ at p_h. If in addition t_2 approaches t_{2h} on the other branch, the adjoint must acquire a node and this node alone takes care of the contacts. Hence the linear system (∞^{n-1}), $H(x^{n+1}, t_{1h}, t_{2h}, t^{n-1}) = 0$, is the system of adjoints nodal at p_h. Consider again the system consisting of the adjoints

$$H(x^{n+1}, t_{1h}, t_{2h}, t_1, t^{n-2}) = 0.$$

According to (6) this system has base points t_{1h}, t_{2h}, t_1 on $\rho_2^{n+3}(t)$ and the further base points, $x(t_{1h}, t_{2h}) = p_h$, $x(t_{1h}, t_1)$, $x(t_{2h}, t_1)$. Thus the three extra intersections at p_h of the adjoints nodal at p_h are accounted for and two outside base points appear. Suppose now that t_1 approaches t_{1h}. The point $x(t_{2h}, t_1)$ approaches the direction at p_h on the tangent t_{2h} of $K(t)$. The point $x(t_{1h}, t_1)$ approaches the point t_{1h} of $K(t)$. Thus we have:

(8) *The linear system (∞^{n-2}) of adjoints $H(x^{n+1}, t_{1h}^2, t_{2h}, t^{n-2}) = 0$ is nodal at p_h with nodal tangents consisting of the tangent to $\rho_2^{n+3}(t)$ at p_h on the branch t_{1h} and of the tangent t_{2h} of $K(t)$. The system has a further base point at the point t_{1h} of $K(t)$.*

Since this linear system (8) has a fixed node at p_h with fixed nodal tangents, it contains a subsystem (∞^{n-3}) with a triple point at p_h which completes the

proof of (7). It is, however, clear that the like linear system (∞^{n-2}) consisting of $H(x^{n+1}, t_{1h}, t_{2h}^2, t^{n-2}) = 0$ also contains the system with triple point p_h, whence

(9) *The linear system of adjoints with a triple point at p_h has the equation $H(x^{n+1}, t_{1h}^2, t_{2h}^2, t^{n-3}) = 0$. This system has base points at the points t_{1h}, t_{2h} of $K(t)$.*

It is perhaps worth noting that, if $\lambda_0, \cdots, \lambda_{n-1}$ are the parameters of the system of adjoints nodal at p_h, the three further conditions, linear in the λ's, that the adjoints have a triple point at p_h yield three primes in the space $[\lambda]$ of a pencil. This is $n - 2$ conditions on the three primes which is precisely the number of conditions on $\rho_2^{n+3}(t)$ that it have a perspective conic. It may be then that, if the property (7) appears at one node of $\rho_2^{n+3}(t)$, it must appear at every other node. This indeed is true in the first pertinent case $n = 3$.

From (7) and (9) we deduce the following theorem toward which the argument has been pointed:

(10) *At any node of $\rho_2^{n+3}(t)$, such as p_1, the lines $(p_1p_2), \cdots, (p_1p_{N_n})$ are projective to the tangents at p_1 to the curves $C_2, \cdots, C_{N_n}$.*

For, the products $(p_1p_2)\cdot C_2, \cdots, (p_1p_{N_n})\cdot C_{N_n}$ are adjoints of order $n + 1$ nodal at p_1. Because of the existence of the system (9) with triple point at p_1, the pairs of nodal tangents are pairs of an involution. This involutorial correspondence has, in the case of the above products, the corresponding pairs mentioned in the theorem.

The property (10) translated to CW_n mapped from $P_{N_n}^2$ states that the points m_{ij} on the line l_i and the primes μ_{ji} on the $[n - 1]$, λ_i are projective, an obvious first requirement if CW_n is to admit a correlation and thus be intrinsically self-dual. We proceed to find this correlation which is a polarity in a quadric Q.

Let the incidence condition of tangent t_1 of $K(t)$ with point t of $\rho_2^{n+3}(t)$ be given, after factoring out (tt_1), by the equation

$$(11) \qquad (k_1t_1)(kt)^{n+2} = 0.$$

This form in cogredient variables t_1, t has precisely the number, $2n + 2$, of absolute constants required to determine uniquely $\rho_2^{n+3}(t)$ with perspective $K(t)$. For variable t_1, (11) determines a pencil of collinear $(n + 2)$-points of $\rho_2^{n+3}(t)$. If t, t' belong to the same member of this pencil, then

$$(12) \qquad Q \equiv (\delta t)^{n+1}(\delta_1t')^{n+1} = (k_1k_1')(kt)^{n+2}(k't')^{n+2}/(tt') = 0.$$

We consider now the mapping of the plane of $\rho_2^{n+3}(t)$ upon the points of a White surface W_n in $[n + 1]$ and the allied mapping of collinear $(n + 1)$-points of $\rho_2^{n+3}(t)$ upon the primes of a White envelope $\overline{W}_n$ developed in the preceding section. The points t_1 of $\rho_2^{n+3}(t)$ map into the norm-curve $N^{n+1}(t)$ on W_n. Then Q in (12) can be interpreted as a quadric Q with apolar point pairs t, t' on $N^{n+1}(t)$. For given t in $Q = 0$ the $n + 1$ points t' are on a prime, the polar prime of t as to Q, and since these $n + 1$ points t' on $\rho_2^{n+3}(t)$ are collinear, this polar prime

is on $\overline{W}_n$. Since the coordinates of the prime are of degree $n + 1$ in t, these polar primes are those of a norm-curve $\bar{N}^{n+1}(t)$ on $\overline{W}_n$. Thus $Q = 0$ is the incidence condition of prime t of $\bar{N}$ with point t' of N. If the node p_h of $\rho_2^{n+3}(t)$ has parameters t_{1h}, t_{2h}, the tangent t_{1h} of $K(t)$ cuts $\rho_2^{n+3}(t)$ again in points t_{2h}, $r_1, \cdots, r_{n+1}$, and the tangent t_{1h} of $K(t)$ cuts in points t_{2h}, $s_1, \cdots, s_{n+1}$. Thus in $Q = 0$ we have, for $t = t_{2h}$, the $n + 1$ values $t' = r_1, \cdots, r_{n+1}$. Now the adjoint of order $n + 1$ on $r_1, \cdots, r_{n+1}$, being on p_h also, must contain as a factor the tangent t_{1h} of $K(t)$, the residual factor being C_h. Hence the corresponding prime of $\bar{N}$ is a prime on the $[n - 1]$, λ_h, with parameter t_{2h}. Thus l_h of CW_n, bisecant to N^{n+1}, corresponds under the polarity Q to λ_h on two primes of $\bar{N}^{n+1}$. Hence

(13) *If $\rho_2^{n+3}(t)$ with nodal $P_{N_n}^2$ has a perspective conic, the nodular set $P_{N_n}'^2$ of* (9) *of* §2 *is projective to $P_{N_n}^2$. The double-N_n configuration mapped from $P_{N_n}^2$ is self-polar under a polarity Q which interchanges the line l_h and the $[n - 1]$, λ_h. Also Q interchanges the W_n on the points of the N_n lines l with $\overline{W}_n$ on the primes of the N_n spaces λ, and the N^{n+1} bisecant to the lines l with the $\bar{N}^{n+1}$ bisecant to the spaces λ.*

This is the type of double-N_n configuration which retains for generic n all the salient features of the double-six on a cubic surface.

5. **The possibility of configurations C_{N_n} more general than CW_n.** At the outset we raised the question as to whether the configurations CW_n are the most general of their type. If we ask for lines l_h and spaces $[n - 1]$, λ_h in $[n + 1]$ such that l_h and λ_h are not incident and that l_h and λ_k $(h \neq k)$ are incident, we would certainly not expect to find only solutions which are "lopsided". This unfortunate lapse characterizes the CW_n. For either the points m_{ij} on l_i, or the primes μ_{ij} on λ_j, define a set $P_{N_n}^2$, but not both, unless conditions are imposed on CW_n. Clearly, then, there is a presumption that more general configurations C_{N_n} exist which are symmetrical to the extent that neither the points m_{ij}, nor the primes μ_{ij}, define a set $P_{N_n}^2$ and that the lopsidedness mentioned is the result of conditions imposed on C_{N_n} to make it a CW_n.

Room [**5**; 376, footnote] expresses the opinion that more general C_{10}'s than those of CW_3 do not exist. He proves (p. 74) that, given in $[n + 1]$ the three lines l_1, l_2, l_3 and $n + 2$ of the $[n - 1]$'s, namely, $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \cdots, \lambda_{n+2}$ with the proper incidences, a unique C_{N_n} can be constructed to contain these spaces and that this is a CW_n. Yet the restriction to *three* lines l_1, l_2, l_3 inevitably suggests a $T(2, n, n + 1)$, which the argument confirms, and one is tempted to begin with a more symmetrical choice of the given elements, i.e., only pairs of opposite elements, l_1, λ_1; l_2, λ_2; $\cdots$.

For cases $n = 1$ and $n = 2$ which are completely known we have the statements:

$n = 1$. Given in [2], 2 opposite pairs with the proper incidences, there is $\infty^0 = 1 C_3$ which contains the two given pairs.

$n = 2$. Given in [3], 3 opposite pairs with the proper incidences, there are $\infty^1 C_6$'s which contain the three given pairs.

Thus one would certainly feel justified in the expectation that

$n = 3$. Given in [4], 4 opposite pairs with the proper incidences, there are $\infty^{k>1} C_{10}$'s which contain the four given pairs.

If this expectation is correct the C_{10}'s would be more general than the CW_3's. For a CW_3 has the 12 absolute constants of a P_{10}^2 whereas the given four opposite lines and planes of a C_{10} with the given incidences already have 12 absolute constants. The most plausible assumption is that $k = 3$, that it imposes three conditions to ask that the primes μ_{ij} define a P_{10}^2, and that it imposes three more to ask that the points m_{ij} define a $P_{10}'^2$. Then the intermediate lopsidedness is removed and P_{10}^2, $P_{10}'^2$ are the nodes of two paired rational sextics.

To see the implications of this conjecture, let us examine the cases $n = 1$ and $n = 2$. The first is obvious, for the given lines l_1, l_2 on points λ_1, λ_2 respectively determine $\lambda_3 = (l_1, l_2)$ and $l_3 = (\lambda_1, \lambda_2)$. In the second case of the double-six of a cubic surface, let the lines l_1, l_2, l_3 and the lines λ_1, λ_2, λ_3 be given with the incidence of l_i and λ_j when $i \neq j$. We wish to determine three lines l_4, l_5, l_6 across λ_1, λ_2, λ_3, and three lines λ_4, λ_5, λ_6 across l_1, l_2, l_3 such that l_k is incident with λ_m, λ_n $(k, m, n = 4, 5, 6)$. Since the given lines involve only 18 constants and the double-six involves 19, we expect ∞^1 solutions. What we wish to emphasize is that *one* solution implies ∞^1 solutions. For, the lines across l_1, l_2, l_3 are those of a regulus R_τ. The coordinates of such a line are quadratic in the parameter τ. Similarly the lines across λ_1, λ_2, λ_3 belong to a regulus R_t and have coordinates quadratic in t. The incidence condition of line τ of R_τ and line t of R_t is the vanishing of a double binary form, $f(t^2, \tau^2) = (a\tau)^2(\alpha t)^2 = 0$. If, then, there is one double-six, this form has one closed configuration, $\tau = \tau_4, \tau_5, \tau_6$, $t = t_4, t_5, t_6$ such that $f(t^2, \tau_k^2) = 0$ has roots t_m, t_n. But the existence of one such closed configuration for f implies the existence of ∞^1 (see [3]). Thus the three given pairs l_i, λ_i define a *poristic form*, $f(t^2, \tau^2)$, and the configurations defined by this poristic form yield the ∞^1 required C_6's.

Naturally this well-known case admits of complete exploration. If (l_i, λ_j) is the plane containing the incident lines l_i, λ_j, there is a pencil of cubic surfaces containing the first three given pairs, namely,

$$(l_1\lambda_2)(l_2\lambda_3)(l_3\lambda_1) + k(l_1\lambda_3)(l_2\lambda_1)(l_3\lambda_2) = 0,$$

the residual cubic curve of the base being composed of the three lines such as $(l_1\lambda_2) = (l_2\lambda_1) = 0$. This pencil cuts R_τ and R_t in projectively related triads of lines,

$$(\gamma t)^3 + k(\delta t)^3 = 0, \qquad (c\tau)^3 + k(d\tau)^3 = 0.$$

Thus we find the obviously poristic determinant form,

$$D_{3,3} = \begin{vmatrix} (\gamma t)^3(\delta t)^3 \\ (c\tau)^3(d\tau)^3 \end{vmatrix} = 0,$$

closed with respect to triads t_4, t_5, t_6; τ_4, τ_5, τ_6, any value of which determines k, and therefore the set of six values. But a generator $t = t_4$ of R_t cuts R_τ in only two points, and therefore fails to meet one generator τ_4 of R_τ. Thus $D_{3,3}$ factors into two forms,

$$D_{3,3} = g(t^1, \tau^1)\cdot f(t^2, \tau^2),$$

the first factor indicating non-incidence, the second, incidence of generators t, τ, and the closure of $D_{3,3}$ implies the closure, or poristic character, of each factor.

Part of the above argument in connection with a C_6 goes on to the C_{10} in a space [4] determined by four given line-plane pairs, l_1, λ_1; $\cdots$; l_4, λ_4. The given pairs have 12 absolute constants so that there must be at least a finite number of even the CW_3's determined by them. Across the four planes $\lambda_1, \cdots, \lambda_4$ there is a set of ∞^2 lines, one line through each generic point of a particular plane. If a, b, c are a fixed triangle on λ_1 and $x_0a + x_1b + x_2c$ a generic point x, the coordinates of the line through x across λ_2, λ_3, λ_4 are cubic polynomials in x which vanish at the points x in which λ_1 is met by λ_2, λ_3, λ_4. Let us call this line l_x. Similarly, if $y = y_0$, y_1, y_2 is a generic prime on l_1, this prime contains a plane λ_y which crosses the four lines l_1, l_2, l_3, l_4. The coordinates of this plane are cubic polynomials in y which vanish for the three positions y determined by the primes containing l_1 and one of l_2, l_3, l_4. The incidence condition of the line l_x and the plane λ_y is $f(x^3, y^3) = 0$. The question we raise is whether this form is *poristic* with infinitely many configurations $x^{(5)}, \cdots, x^{(10)}$; $y^{(5)}, \cdots, y^{(10)}$ such that $f(x^{(k)^3}, y^{(l)^3}) = 0$ for $k, l = 5, \cdots, 10$; $k \neq l$. Evidently each such configuration gives rise to six pairs l_5, λ_5; $\cdots$; l_{10}, λ_{10} of line-planes which extend the four given pairs to make up a configuration C_{10}, and a poristic form would yield configurations which are not CW_3's. Since there is as yet no theory for the poristic double ternary form, we leave the problem at this point.

Bibliography

1. A. B. Coble, *Algebraic Geometry and Theta Functions*, American Mathematical Society Colloquium Publications, vol. 10(1929).
2. A. B. Coble, *Conditions on the nodes of a rational plane curve*, this Journal, vol. 7(1940), pp. 396–410.
3. A. B. Coble, *Multiple binary forms with the closure property*, American Journal of Mathematics, vol. 43(1921), pp. 1–19.
4. A. B. Coble, *Trilinear forms*, this Journal, vol. 7(1940), pp. 380–395.
5. T. G. Room, *The Geometry of Determinantal Loci*, Cambridge University Press, London, 1938.
6. F. Schur, *Ueber die durch collineare Grundgebilde erzeugten Curven und Flächen*, Mathematische Annalen, vol. 18(1881), pp. 1–32.
7. F. P. White, *On certain nets of plane curves*, Proceedings of the Cambridge Philosophical Society, vol. 22(1923), pp. 1–10.

University of Illinois.

A PARTICULAR SET OF TEN POINTS IN SPACE

By Arthur B. Coble

1. **Introduction.** A generic trilinear form $T(2, 3, 4) = (\alpha x)(\beta y)(\gamma z)$ with digredient variables x, y, z which are contragredient to ξ, η, ζ respectively in their spaces [2], [3], [4] depends upon 12 absolute constants. There are 10 pairs $x, y = p_i, q_i$ $(i = 1, \cdots, 10)$ which are neutral for z in $T = 0$. In an earlier paper [4], it is proved that the set of ten points p_i, P_{10}^2 and the set of ten points q_i, Q_{10}^3 are connected by the double identity in ξ, η,

$$\sum_{i=1}^{i=10} (q_i \eta) \cdot (p_i \xi)^2 \equiv 0. \tag{1}$$

In this identity the set P_{10}^2 is generic with 12 absolute constants. The set Q_{10}^3 is then projectively determined by the identity and thus is subject to three projective conditions. It is the purpose of this paper to determine the nature of these conditions, and so explore some of their consequences.

2. **The generic character of 9 points of Q_{10}^3.** In this section we prove that the three conditions on Q_{10}^3 all fall on the tenth point when the first nine are given generically. This is somewhat unusual. For example, the ten nodes of a rational sextic are subject to three conditions and only eight can be chosen generically; in space the nine nodes of a symmetroid are subject to three conditions and only seven can be chosen generically. We observe first that the squares $(p\xi)^2$ of points p in [2] represent a mapping of the plane [2] upon the points r of a Veronese V_2^4 in [5], the point p_i mapping into a point r_i. Thus we have a set R_{10}^5 on V_2^4, and the identity (1) asserts that the set Q_{10}^3 is associated to the set R_{10}^5. Hence

(1) *The three conditions on Q_{10}^3 appear in its associated set R_{10}^5 as the three conditions that R_{10}^5 is on a Veronese V_2^4.*

For, it is known [**3**; Theorem 18] that on nine generic points in [5] there are four V_2^4's, whence the three conditions on R_{10}^5 fall on the tenth when the first nine are given. Let then $r_1, \cdots, r_9$ be projected from r_{10} into a set $S_9^4 = s_1, \cdots, s_9$ in [4], the V_2^4 on R_{10}^5 projecting into a M_2^3 on S_9^4. This S_9^4 is associated to $Q_9^3 = q_1, \cdots, q_9$. Again it is known [**3**; Theorem 15] that on an S_9^4 there are two M_2^3's, these being paired with the two reguli on the associated Q_9^3. Since any M_2^3 in [4] is the map of the plane by conics on a point, any S_9^4 and M_2^3 on it can be obtained by such a mapping from nine points of the plane. Since the above S_9^4 is obtained from $p_1, \cdots, p_9$ by the mapping with conics on p_{10}, the S_9^4 is a generic set, and its associated set Q_9^3 is also generic. Hence

Received March 24, 1942.

(2) *Any nine points of Q^3_{10} constitute a generic set of nine points, and the two reguli on the nine points are separated.*

3. **The construction of Q^3_{10} when nine points are given.** Let $x^{(0)}$, $z^{(0)}$ be neutral for y in $T = 0$, i.e., $(\alpha x^{(0)})(\gamma z^{(0)})\beta_j = 0$. Then, for given $x^{(0)}$ in [2], $z^{(0)}$ is on the Bordiga [**1**] surface F^6_2 in [4], the White surface W_3 of the preceding paper. For given z in [4], there is a point y in [3] which with z is neutral for x in $T = 0$, i.e., $\alpha_i(\gamma z)(\beta y) = 0$. However, for a point $z^{(0)}$ on F^6_2, these equations in y are linearly related, and y is on a line $\rho^{(0)}$ of the Semple [**7**] congruence. The properties of this correspondence between $x^{(0)}$ in [2], $z^{(0)}$ on F^6_2 in [4], and line $\rho^{(0)}$ in [3] have been summarized in Room [6], and we apply them without derivation. To the directions p^*_i about p_i, there correspond the points $z^{(0)}$ on a line l_i of F^6_2, and the lines $\rho^{(0)}$ of a regulus $R_i(t)$ of the quadric B_i on all of the points of Q^3_{10} except q_i. This immediately indicates a separation of the reguli on all ten of the quadrics B_i into reguli $R_i(t)$ and $R_i(\tau)$. Again to the points $x^{(0)}$ of the cubic curve C_j on all the points of P^2_{10} except p_j there correspond the points $z^{(0)}$ of a cubic curve k_j on F^6_2 in a plane λ_j, and the generators $\rho^{(0)}$ of a cubic cone K_j with vertex at q_j and on the other points of Q^3_{10}.

From the obvious fact that the cubic curve C_j has a direction at p_i $(i \neq j)$ we conclude that

(1) *The cubic cone K_j contains that generator of the regulus $R_i(t)$ which passes through q_j.*

From this and (2) of the preceding section, we obtain at once the following construction for Q^3_{10}.

(2) *Given nine generic points $q_1, \cdots, q_9$ in* [3] *and the quadric B_{10} on them with reguli $R_{10}(t)$, $R_{10}(\tau)$, the cubic cones K_j $(j = 1, \cdots, 9)$, with triple point at q_j, on the remaining eight given points, and on the generator of $R_{10}(t)$ through q_j, all pass through the point q_{10} of Q^3_{10}. The cones K'_j, similarly defined for the regulus $R_{10}(\tau)$ of B_{10} determine a point q'_{10} which with $q_1, \cdots, q_9$ make up a set Q'^3_{10}.*

These theorems, (1) and (2), indicate how the separation of the reguli on the quadrics $B_1, \cdots, B_{10}$ of Q^3_{10} is effected.

4. **A comparison of Q^3_{10} with Σ^3_{10}.** It is clear from the preceding paper that the set Q^3_{10} under consideration, subject to three conditions, is half way along the path of specialization to the set Σ^3_{10} of nodes of a symmetroid, a set which is subject to six conditions. Some properties of Q^3_{10} associated with the system (∞^4) of Reye sextics on Q^3_{10} are found in §§5–8 of the article cited in [**4**]. We shall be concerned here more particularly with properties which it shares in whole or part with Σ^3_{10}. One noteworthy property [**5**; 39] of Σ^3_{10} is that if any one of the quadrics $B_1, \cdots, B_{10}$ has a node, each of the others also has a node.

Thus the discriminants $\Delta_1, \cdots, \Delta_{10}$ of $B_1, \cdots, B_{10}$ are identical. In contrast with this we have

(1) *If Q_{10}^3 is given, the radicals $(\Delta_1)^{\frac{1}{2}}, \cdots, (\Delta_{10})^{\frac{1}{2}}$ each are rational functions of any one of them, and of the given Q_{10}^3.*

This is a consequence of the construction (2) of the preceding section. For, if a value of $(\Delta_{10})^{\frac{1}{2}}$ is given to isolate the regulus $R_{10}(t)$ on B_{10} and thus to isolate q_{10} from q_{10}', there is equally well isolated a regulus $R_i(t)$ on B_i $(i = 1, \cdots, 9)$, and thus a value of $(\Delta_i)^{\frac{1}{2}}$.

There is, however, a considerable difference in the application of the construction (2) of §3 to Σ_{10}^3 and to Q_{10}^3. If $q_1, \cdots, q_9$ are subject to the three conditions that they belong to Σ_{10}^3, there is a *pencil* of cones K_j and there is necessarily a member of this pencil on each of the generators of B_{10} through q_j. Thus q_{10} and q_{10}' *coincide* (three conditions). If, however, $q_1, \cdots, q_9$ of Q_{10}^3 are subject to the one condition that B_{10} have a node, and thus the reguli $R_{10}(t)$, $R_{10}(\tau)$ coincide, then q_{10}' approaches q_{10} in some direction (one condition). It is not then necessarily true that q_1' approaches coincidence with q_1 for the given $q_2, \cdots, q_{10}$.

Another interesting relation of the two sets Q_{10}^3 of (2) of §3 is the following:

(2) *If $Q_{10}^3 = q_1, \cdots, q_9, q_0$ and $Q_{10}'^3 = q_1, \cdots, q_9, q_0'$ are constructed as in* (2) *of* §3, *and if $p_1, \cdots, p_9, p_0$ and $p_1', \cdots, p_9', p_0'$ are the ternary sets P_{10}^2, $P_{10}'^2$ related to them as in* (1) *of* §1, *then P_{10}^2, $P_{10}'^2$ are the direct and inverse F-points of a Cremona transformation R_0 of order* 17 *with* 12*-fold F-points at p_0, p_0' and* 4*-fold F-points at p_i, p_i' $(i = 1, \cdots, 9)$.*

Under this transformation R_0, the directions at p_0 correspond to the P-curve defined at $P_{10}'^2$ by $(0^8 i^3)^{12}$, and the directions at p_i to the P-curve $(0^3 i^0 j^1)^4$ $(j \neq i)$ [2]. We recall from §2 that $q_1, \cdots, q_9$ are associated to $S_9^4 = s_1, \cdots, s_9$ in [4] and that the plane of P_{10}^2 is mapped by conics on p_0 into an M_2^3 on S_9^4, the points $p_1, \cdots, p_9$ mapping into $s_1, \cdots, s_9$. If $P_{10}'^2$ is the set of inverse F-points of R_0 with F-points at P_{10}^2, and if also $p_{i_1}, \cdots, p_{i_5}$ are on a conic with p_0, then $p_{i_1}', \cdots, p_{i_5}'$ are on a conic with p_0' $(i_1, \cdots, i_5 = 1, \cdots, 9)$. Hence, conics on p_0' map the plane of $P_{10}'^2$ into an $M_2'^3$ on $S_9'^4$ with S_9^4 and $S_9'^4$ so related that if five points of S_9^4 are on a prime, the like-named points of $S_9'^4$ are also on a prime, and vice versa. Hence, S_9^4 and $S_9'^4$ are projective and may be superposed. Since lines on p_0, which map into generators of M_2^3, do not pass by R_0 into lines on p_0', the $M_2'^3$ on $S_9'^4$ is not superposed on the M_2^3 on S_9^4. Thus the two planar sets are the two projectively distinct sets which determine Q_{10}^3, $Q_{10}'^3$ respectively.

In case Q_{10}^3 is a Σ_{10}^3, the two sets Q_{10}^3, $Q_{10}'^3$ coincide throughout. However, there are still two projectively distinct sets P_{10}^2, $P_{10}'^2$ related as in the theorem, and related as above to the two M_2^3's on S_9^4, or to the two reguli on the quadric B_0. These are now the two sets of nodes of two "paired" rational sextics [**2**; 252, (4)].

Conditions on P_{10}^2, $P_{10}'^2$ equivalent to those given in (2) can be expressed in the following simpler form.

(3) *The quadratic transformation $A_{0i_1i_2}$ with F-points at p_0, p_{i_1}, p_{i_2} transforms the remaining seven points of P_{10}^2 into a set R_7^2 projective to the set $R_7'^2$ into which the remaining seven points of $P_{10}'^2$ are carried by $A_{0i_1i_2}'$.*

For, the projection of S_9^4 from s_{i_1}, s_{i_2} yields a set R_7^2 which is associated to the set Q_7^3 obtained from $q_1, \cdots, q_9$ by deleting q_{i_1}, q_{i_2}. But two sets R_7^2, $R_7'^2$ each associated to Q_7^3 are projective to each other. This theorem (3) also applies to the case of the nodes of two paired rational sextics and then is equivalent to the statement given elsewhere [**2**; 253, (7)].

We prove finally that the conditions on Q_{10}^3 are invariant under properly chosen Cremona transformation, the precise statement being:

(4) *If Q_{10}^3 is related to P_{10}^2 as in* (1) *of* §1, *and if Q_{10}^3 is congruent to $Q_{10}'^3$ under the regular cubic transformation A_{1234}, and P_{10}^2 is congruent to $P_{10}'^2$ under the quintic transformation A_{567890}, then $Q_{10}'^3$ is still related to $P_{10}'^2$ as in* (1) *of* §1.

For, if $P_{10}'^2$ is congruent to P_{10}^2 under A_{567890}, the set R_{10}^5 on V_2^4 obtained from P_{10}^2 is congruent to a set $R_{10}'^5$ on $V_2'^4$ obtained from R_{10}^5 by the regular transformation in [5] with F-points at $r_5, \cdots, r_9$, r_0 [**3**; 16]. Then their associated sets are congruent under A_{1234}.

This is the theorem [**2**; 257, (14)] concerning $Q_{10}^3 = \Sigma_{10}^3$ and P_{10}^2 = nodes of $\rho_2^6(t)$. The difference in the two cases is that the number of projectively distinct congruent sets is finite in the case of Σ_{10}^3, and infinite in the case of Q_{10}^3.

Bibliography

1. G. Bordiga, *La superficie del 6° ordine con 10 rette, nello spazio* R$_4$ *e le sue projezione nello spazio ordinario*, Atti R. Acad. Lincei (Mem.), vol. 4(1887), pp. 182–203.
2. A. B. Coble, *Algebraic Geometry and Theta Functions*, Colloquium Publications, American Mathematical Society, vol. 10(1929), pp. 252–254.
3. A. B. Coble, *Associated sets of points*, Transactions of the American Mathematical Society, vol. 24(1922), pp. 1–20.
4. A. B. Coble, *Trilinear forms*, this Journal, vol. 7(1940), pp. 380–395.
5. J. R. Conner, *The rational sextic curve, and the Cayley symmetroid*, American Journal of Mathematics, vol. 37(1915), pp. 29–42.
6. T. G. Room, *The Geometry of Determinantal Loci*, Cambridge University Press, London, 1938, pp. 384–385.
7. J. G. Semple, *On representations of line-congruences of the second and third orders*, Proceedings of the London Mathematical Society (2), vol. 35(1933), pp. 294–324.

University of Illinois.

CONTENTS

CONSERVATION OF SCHOLARLY JOURNALS

The American Library Association created this last year the Committee on Aid to Libraries in War Areas, headed by John R. Russell, the Librarian of the University of Rochester. The Committee is faced with numerous serious problems and hopes that American scholars and scientists will be of considerable aid in the solution of one of these problems.

One of the most difficult tasks in library reconstruction after the first World War was that of completing foreign institutional sets of American scholarly, scientific, and technical periodicals. The attempt to avoid a duplication of that situation is now the concern of the Committee.

Many sets of journals will be broken by the financial inability of the institutions to renew subscriptions. As far as possible they will be completed from a stock of periodicals being purchased by the Committee. Many more will have been broken through mail difficulties and loss of shipments, while still other sets will have disappeared in the destruction of libraries. The size of the eventual demand is impossible to estimate, but requests received by the Committee already give evidence that it will be enormous.

With an imminent paper shortage attempts are being made to collect old periodicals for pulp. Fearing this possible reduction in the already limited supply of scholarly and scientific journals, the Committee hopes to enlist the cooperation of subscribers to this journal in preventing the sacrifice of this type of material to the pulp demand. It is scarcely necessary to mention the appreciation of foreign institutions and scholars for this activity.

Questions concerning the project or concerning the value of particular periodicals to the project should be directed to Wayne M. Hartwell, Executive Assistant to the Committee on Aid to Libraries in War Areas, Rush Rhees Library, University of Rochester, Rochester, New York.

(FOR CONTENTS, SEE INSIDE BACK COVER.)

on
the
ous
ible

irst
ool-
tion

in-
ted
will
still
of
the

old
ited
the
this
the

ular
tive
nces